SUBSTITUTE SEQUENCE LISTING

<110>    CHIRON CORPORATION
         HARDY, Stephen F
         DONNELLY, III, John J
         ZUR MEGEDE, Jan T

<120>    VECTORS FOR EXPRESSION OF HML-2 POLYPEPTIDES

<130>    PP19482.0007


<140>    10/587,032
<141>    2006-07-24

<150>    PCT/US03/18666
<151>    2003-06-13

<150>    60/388831
<151>    2002-06-16

<150>    60/472189
<151>    2003-05-20

<160>    83

<170>    PatentIn, version 3.5

<210>    1
<211>    1998
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    1
atggggcaaa ctaaaagtaa aattaaaagt aaatatgcct cttatctcag ctttattaaa    60
attcttttaa aaagaggggg agttaaagta tctacaaaaa atctaatcaa gctatttcaa    120
ataatagaac aattttgccc atggtttcca gaacaaggaa ctttagatct aaaagattgg    180
aaaagaattg gtaaggaact aaaacaagca ggtaggaagg gtaatatcat tccacttaca    240
gtatggaatg attgggccat tattaaagca gctttagaac catttcaaac agaagaagat    300
agcgtttcag tttctgatgc ccctggaagc tgtataatag attgtaatga aaacacaagg    360
aaaaaatccc agaaagaaac ggaaggttta cattgcgaat atgtagcaga gccggtaatg    420
gctcagtcaa cgcaaaatgt tgactataat caattacagg aggtgatata tcctgaaacg    480
ttaaaattag aaggaaaagg tccagaatta gtggggccat cagagtctaa accacgaggc    540
acaagtcctc ttccagcagg tcaggtgcct gtaacattac aacctcaaaa gcaggttaaa    600
gaaaataaga cccaaccgcc agtagcctat caatactggc ctccggctga acttcagtat    660
cggccacccc cagaaagtca gtatggatat ccaggaatgc ccccagcacc acagggcagg    720
gcgccatacc ctcagccgcc cactaggaga cttaatccta cggcaccacc tagtagacag    780
ggtagtaaat tacatgaaat tattgataaa tcaagaaagg aaggagatac tgaggcatgg    840
caattcccag taacgttaga accgatgcca cctggagaag gagcccaaga gggagagcct    900
cccacagttg aggccagata caagtctttt tcgataaaaa agctaaaaga tatgaaagag    960
ggagtaaaac agtatggacc caactcccct tatatgagga cattattaga ttccattgct    1020
catggacata gactcattcc ttatgattgg gagattctgg caaaatcgtc tctctcaccc    1080
tctcaatttt acaatttaa gacttggtgg attgatgggg tacaagaaca ggtccgaaga    1140
aatagggctg ccaatcctcc agttaacata gatgcagatc aactattagg aataggtcaa    1200
aattggagta ctattagtca caagcatta atgcaaaatg aggccattga gcaagttaga    1260
gctatctgcc ttagagcctg ggaaaaaatc caagacccag gaagtacctg cccctcattt    1320
aatacagtaa gacaaggttc aaaagagccc tatcctgatt ttgtggcaag gctccaagat    1380
gttgctcaaa agtcaattgc tgatgaaaaa gcccgtaagg tcatagtgga gttgatggca    1440
tatgaaaacg ccaatcctga gtgtcaatca gccattaagc cattaaaagg aaaggttcct    1500
gcaggatcag atgtaatctc agaatatgta aaagcctgtg atggaatcgg aggagctatg    1560
cataaagcta tgcttatggc tcaagcaata acaggagttg ttttaggagg acaagttaga    1620
acatttggaa gaaaatgtta taattgtggt caaattggtc acttaaaaaa gaattgccca    1680
gtcttaaata aacagaatat aactattcaa gcaactacaa caggtagaga gccacctgac    1740

Page 1

```
ttatgtccaa gatgtaaaaa aggaaaacat tgggctagtc aatgtcgttc taaatttgat   1800
aaaaatgggc aaccattgtc gggaaacgag caaaggggcc agcctcaggc cccacaacaa   1860
actggggcat tcccaattca gccatttgtt cctcagggtt ttcagggaca acaaccccca   1920
ctgtcccaag tgtttcaggg aataagccag ttaccacaat acaacaattg tcccccgcca   1980
caagcggcag tgcagcag                                                 1998
```

<210>    2
<211>    2001
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    2
```
atggggcaaa ctaaaagtaa aattaaaagt aaatatgcct cttatctcag ctttattaaa   60
attctttaa aaagagggggg agttaaagta tctacaaaaa atctaatcaa gctatttcaa   120
ataatagaac aattttgccc atggtttcca gaacaaggaa ctttagatct aaaagattgg   180
aaaagaattg gtaaggaact aaaacaagca ggtaggaagg gtaatatcat tccacttaca   240
gtatggaatg attgggccat tattaaagca gctttagaac catttcaaac agaagaagat   300
agcgtttcag tttctgatgc ccctggaagc tgtataatag attgtaatga aaacacaagg   360
aaaaaatccc agaaagaaac ggaaggttta cattgcgaat atgtagcaga gccggtaatg   420
gctcagtcaa cgcaaaatgt tgactataat caattacagg aggtgatata tcctgaaacg   480
ttaaaattag aaggaaaagg tccagaatta gtggggccat cagagtctaa accacgaggc   540
acaagtcctc ttccagcagg tcaggtgcct gtaacattac aacctcaaaa gcaggttaaa   600
gaaaataaga cccaaccgcc agtagcctat caatactggc ctccggctga acttcagtat   660
cggccacccc cagaaagtca gtatggatat ccaggaatgc ccccagcacc acagggcagg   720
gcgccatacc ctcagccgcc cactaggaga cttaatccta cggcaccacc tagtagacag   780
ggtagtaaat tacatgaaat tattgataaa tcaagaaagg aaggagatac tgaggcatgg   840
caattcccag taacgttaga accgatgcca cctggagaag gagcccaaga gggagagcct   900
cccacagttg aggccagata caagtctttt tcgataaaaa agctgaaaga tatgaaagag   960
ggagtaaaac agtatggacc caactcccct tatatgagga cattattaga ttccattgct   1020
catggacata gactcattcc ttatgattgg gagattctgg caaaatcgtc tctctcaccc   1080
tctcaatttt tacaatttaa gacttggtgg attgatgggg tacaagaaca ggtccgaaga   1140
aatagggctg ccaatcctcc agttaacata gatgcagatc aactattagg aataggtcaa   1200
aattggagta ctattagtca acaagcatta atgcaaaatg aggccattga gcaagttaga   1260
gctatctgcc ttagagcctg ggaaaaaatc caagacccag gaagtacctg cccctcattt   1320
aatacagtaa gacaaggttc aaaagagccc tatcctgatt ttgtggcaag gctccaagat   1380
gttgctcaaa agtcaattgc tgatgaaaaa gcccgtaagg tcatagtgga gttgatggca   1440
tatgaaaacg ccaatcctga gtgtcaatca gccattaagc cattaaaagg aaaggttcct   1500
gcaggatcag atgtaatctc agaatatgta aaagcctgtg atggaatcgg aggagctatg   1560
tataaagcta tgcttatggc tcaagcaata acaggagttg ttttaggagg acaagttaga   1620
acatttggaa gaaatgtta .taattgtggt caaattggtc acttaaaaaa gaattgccca   1680
gtcttaaata aacagaatat aactattcaa gcaactacaa caggtagaga gccacctgac   1740
ttatgtccaa gatgtaaaaa aggaaaacat tgggctagtc aatgtcgttc taaatttgat   1800
aaaaatgggc aaccattgtc gggaaacgag caaaggggcc agcctcaggc cccacaacaa   1860
actggggcat tcccaattca gccatttgtt cctcagggtt ttcagggaca acaaccccca   1920
ctgtcccaag tgtttcaggg aataagccag ttaccacaat acaacaattg tcccccgcca   1980
caagcggcag tgcagcagta g                                             2001
```

<210>    3
<211>    2004
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    3
```
atggggcaaa ctaaaagtaa aactaaaagt aaatatgcct cttatctcag ctttattaaa   60
attctttaa aaagagggggg agttagagta tctacaaaaa atctaatcaa gctatttcaa   120
ataatagaac aattttgccc atggtttcca gaacaaggaa ctttagatct aaaagattgg   180
aaaagaattg gcgaggaact aaaacaagca ggtagaaagg gtaatatcat tccacttaca   240
gtatggaatg attgggccat tattaaagca gctttagaac catttcaaac aaaagaagat   300
agcgtttcag tttctgatgc ccctggaagc tgtgtaatag attgtaatga aaagacaggg   360
agaaaatccc agaaagaaac agaaagttta cattgcgaat atgtaacaga gccagtaatg   420
gctcagtcaa cgcaaaatgt tgactataat caattacagg gggtgatata tcctgaaacg   480
ttaaaattag aaggaaaagg tccagaatta gtggggccat cagagtctaa accacgaggg   540
ccaagtcctc ttccagcagg tcaggtgccc gtaacattac aacctcaaac gcaggttaaa   600
```
                              Page 2

```
gaaaataaga cccaaccgcc agtagcttat caatactggc cgccggctga acttcagtat    660
ctgccacccc cagaaagtca gtatggatat ccaggaatgc ccccagcact acagggcagg    720
gcgccatatc ctcagccgcc cactgtgaga cttaatccta cagcatcacg tagtggacaa    780
ggtggtacac tgcacgcagt cattgatgaa gccagaaaac agggagatct tgaggcatgg    840
cggttcctgg taattttaca actggtacag gccggggaag agactcaagt aggagcgcct    900
gcccgagctg agactagatg tgaacctttc accatgaaaa tgttaaaaga tataaaggaa    960
ggagttaaac aatatggatc caactcccct tatataagaa cattattaga ttccattgct   1020
catggaaata gacttactcc ttatgactgg gaaagtttgg ccaaatcttc cctttcatcc   1080
tctcagtatc tacagtttaa aacctggtgg attgatggag tacaagaaca ggtacgaaaa   1140
aatcaggcta ctaagcccac tgttaatata gacgcagacc aattgttagg aacaggtcca   1200
aattggagca ccattaacca acaatcagtg atgcagaatg aggctattga acaagtaagg   1260
gctatttgcc tcagggcctg gggaaaaatt caggacccag aacagctttc cctattaat    1320
tcaattagac aaggctctaa agagccatat cctgactttg tggcaagatt acaagatgct   1380
gctcaaaagt ctattacaga tgacaatgcc cgaaaagtta ttgtagaatt aatggcctat   1440
gaaaatgcaa atccagaatg tcagtcggcc ataaagccat taaaaggaaa agttccagca   1500
ggagttgatg taattacaga atatgtgaag gcttgtgatg ggattggagg agctatgcat   1560
aaggcaatgc taatggctca agcaatgagg gggctcactc taggaggaca agttagaaca   1620
tttgggaaaa aatgttataa ttgtggtcaa atcggtcatc tgaaaaggag ttgcccagtc   1680
ttaaataaac agaatataat aaatcaagct attacagcaa aaaataaaaa gccatctggc   1740
ctgtgtccaa aatgtggaaa aggaaaacat tgggccaatc aatgtcattc taaatttgat   1800
aaagatgggc aaccattgtc gggaaacagg aagaggggcc agcctcaggc cccccaacaa   1860
actggggcat tcccagttca actgtttgtt cctcagggtt ttcaaggaca acaacccta   1920
cagaaaatac caccacttca gggagtcagc caattacaac aatccaacag ctgtcccgcg   1980
ccacagcagg cagcgccaca gtag                                        2004
```

```
<210>    4
<211>    852
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)
```

```
<400>    4
atggggcaaa ctaaaagtaa aattaaaagt aaatatgcct cttatctcag ctttattaaa     60
attctttaa aaagaggggg agttaaagta tctacaaaaa atctaatcaa gctatttcaa    120
ataatagaac aattttgccc atggtttcca gaacaaggaa cttcagatct aaaagattgg    180
aaaagaattg gtaaggaact aaaacaagca ggtaggaagg gtaatatcat tccacttaca    240
gtatggaatg attgggccat tattaaagca gctttagaac catttcaaac agaagaagat    300
agcatttcag tttctgatgc ccctggaagc tgtttaatag attgtaatga aaacacaagg    360
aaaaaatccc agaaagaaac cgaaagttta cattgcgaat atgtagcaga gccggtaatg    420
gctcagtcaa cgcaaaatgt tgactataat caattacagg aggtgatata tcctgaaacg    480
ttaaaattag aaggaaaagg tccagaatta atggggccat cagagtctaa accacgaggc    540
acaagtcctc ttccagcagg tcaggtgctc gtaagattac aacctcaaaa gcaggttaaa    600
gaaaataaga cccaaccgca agtagcctat caatactgcc gctggctgaa cttcagtatc    660
ggccaccccc agaaagtcag tatggatatc caggaatgcc cccagcacca cagggcaggg    720
cgccatacca tcagccgccc actaggagac ttaatcctat ggcaccacct agtagacagg    780
gtagtgaatt acatgaaatt attgataaat caagaaggaa ggagatact gaggcatggc     840
aattcccagt aa                                                      852
```

```
<210>    5
<211>    666
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)
```

```
<400>    5
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1                5                  10                  15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
            20                  25                  30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
        35                  40                  45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
```

```
          50                      55                      60
```

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65                  70                  75                  80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
                85                  90                  95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
                100                 105                 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
                115                 120                 125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
        130                 135                 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145                 150                 155                 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
                165                 170                 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
                180                 185                 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
        195                 200                 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
        210                 215                 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225                 230                 235                 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
                245                 250                 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
                260                 265                 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
        275                 280                 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
        290                 295                 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305                 310                 315                 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
                325                 330                 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
                340                 345                 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
        355                 360                 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
        370                 375                 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln

385                390                395                400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
            405                410                415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
            420                425                430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
            435                440                445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
        450                455                460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465                470                475                480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
            485                490                495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
            500                505                510

Cys Asp Gly Ile Gly Gly Ala Met Tyr Lys Ala Met Leu Met Ala Gln
            515                520                525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
        530                535                540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545                550                555                560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
            565                570                575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
            580                585                590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
            595                600                605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
    610                615                620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625                630                635                640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
            645                650                655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
            660                665

```
<210>    6
<211>    667
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    6
Met Gly Gln Thr Lys Ser Lys Thr Lys Ser Lys Tyr Ala Ser Tyr Leu
1                5                10                15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Arg Val Ser Thr
```

```
                      20                  25                  30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
            35                  40                  45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
        50                  55                  60

Glu Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65                  70                  75                  80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
                85                  90                  95

Thr Lys Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Val
            100                 105                 110

Ile Asp Cys Asn Glu Lys Thr Gly Arg Lys Ser Gln Lys Glu Thr Glu
        115                 120                 125

Ser Leu His Cys Glu Tyr Val Thr Glu Pro Val Met Ala Gln Ser Thr
    130                 135                 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Gly Val Ile Tyr Pro Glu Thr
145                 150                 155                 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
            165                 170                 175

Lys Pro Arg Gly Pro Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
            180                 185                 190

Leu Gln Pro Gln Thr Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
            195                 200                 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Leu Pro Pro Pro
    210                 215                 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Leu Gln Gly Arg
225                 230                 235                 240

Ala Pro Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
            245                 250                 255

Arg Ser Gly Gln Gly Gly Thr Leu His Ala Val Ile Asp Glu Ala Arg
            260                 265                 270

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
    275                 280                 285

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
    290                 295                 300

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
305                 310                 315                 320

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
            325                 330                 335

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ser
    340                 345                 350

Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
```

```
            355                        360                        365
Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
    370                375                380

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
385                390                395                400

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
            405                410                415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
            420                425                430

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
            435                440                445

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
    450                455                460

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
465                470                475                480

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
            485                490                495

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
            500                505                510

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
            515                520                525

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
    530                535                540

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
545                550                555                560

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
            565                570                575

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
            580                585                590

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
    595                600                605

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
    610                615                620

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
625                630                635                640

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
            645                650                655

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
            660                665
```

```
<210>   7
<211>   283
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)
```

<400> 7

```
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1               5               10              15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
            20              25              30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
        35              40              45

Phe Pro Glu Gln Gly Thr Ser Asp Leu Lys Asp Trp Lys Arg Ile Gly
    50              55              60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65              70              75              80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
            85              90              95

Thr Glu Glu Asp Ser Ile Ser Val Ser Asp Ala Pro Gly Ser Cys Leu
            100             105             110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
        115             120             125

Ser Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
    130             135             140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145             150             155             160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Met Gly Pro Ser Glu Ser
            165             170             175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Leu Val Arg
            180             185             190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Gln Val
            195             200             205

Ala Tyr Gln Tyr Cys Arg Trp Leu Asn Phe Ser Ile Gly His Pro Gln
    210             215             220

Lys Val Ser Met Asp Ile Gln Glu Cys Pro Gln His His Arg Ala Gly
225             230             235             240

Arg His Thr Ile Ser Arg Pro Leu Gly Asp Leu Ile Leu Trp His His
            245             250             255

Leu Val Asp Arg Val Val Asn Tyr Met Lys Leu Leu Ile Asn Gln Glu
            260             265             270

Arg Lys Glu Ile Leu Arg His Gly Asn Ser Gln
        275             280
```

<210> 8
<211> 434
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 8

```
Met Pro Pro Ala Pro Gln Gly Arg Ala Pro Tyr His Gln Pro Pro Thr
```

```
1                      5                    10                        15

Arg Arg Leu Asn Pro Met Ala Pro Pro Ser Arg Gln Gly Ser Glu Leu
            20              25               30

His Glu Ile Ile Asp Lys Ser Arg Lys Glu Gly Asp Thr Glu Ala Trp
            35              40               45

Gln Phe Pro Val Thr Leu Glu Pro Met Pro Pro Gly Glu Gly Ala Gln
    50              55               60

Glu Gly Glu Pro Pro Thr Val Glu Ala Arg Tyr Lys Ser Phe Ser Ile
65                  70               75                        80

Lys Met Leu Lys Asp Met Lys Glu Gly Val Lys Gln Tyr Gly Pro Asn
                85              90               95

Ser Pro Tyr Met Arg Thr Leu Leu Asp Ser Ile Ala Tyr Gly His Arg
            100             105              110

Leu Ile Pro Tyr Asp Trp Glu Ile Leu Ala Lys Ser Ser Leu Ser Pro
            115             120              125

Ser Gln Phe Leu Gln Phe Lys Thr Trp Trp Ile Asp Gly Val Gln Glu
    130             135              140

Gln Val Arg Arg Asn Arg Ala Ala Asn Pro Pro Val Asn Ile Asp Ala
145             150             155                        160

Asp Gln Leu Leu Gly Ile Gly Gln Asn Trp Ser Thr Ile Ser Gln Gln
            165             170              175

Ala Leu Met Gln Asn Glu Ala Ile Glu Gln Val Arg Ala Ile Cys Leu
            180             185              190

Arg Ala Trp Glu Lys Ile Gln Asp Pro Gly Ser Thr Cys Pro Ser Phe
    195             200              205

Asn Thr Val Arg Gln Gly Ser Lys Glu Pro Tyr Pro Asp Phe Val Ala
    210             215              220

Arg Leu Gln Asp Val Ala Gln Lys Ser Ile Ala Asp Glu Lys Ala Gly
225             230             235                        240

Lys Val Ile Val Glu Leu Met Ala Tyr Glu Asn Ala Asn Pro Glu Cys
            245             250              255

Gln Ser Ala Ile Lys Pro Leu Lys Gly Lys Val Pro Ala Gly Ser Asp
            260             265              270

Val Ile Ser Glu Tyr Val Lys Ala Cys Asp Gly Ile Gly Gly Ala Met
    275             280              285

His Lys Ala Met Leu Met Ala Gln Ala Ile Thr Gly Val Val Leu Gly
290             295              300

Gly Gln Val Arg Thr Phe Gly Gly Lys Cys Tyr Asn Cys Gly Gln Ile
305             310              315                        320

Gly His Leu Lys Lys Asn Cys Pro Val Leu Asn Lys Gln Asn Ile Thr
            325             330              335

Ile Gln Ala Thr Thr Thr Gly Arg Glu Pro Pro Asp Leu Cys Pro Arg
```

```
                    340                 345                 350
Cys Lys Lys Gly Lys His Trp Ala Ser Gln Cys Arg Ser Lys Phe Asp
        355             360             365

Lys Asn Gly Gln Pro Leu Ser Gly Asn Glu Gln Arg Gly Gln Pro Gln
    370             375             380

Ala Pro Gln Gln Thr Gly Ala Phe Pro Ile Gln Pro Phe Val Pro Gln
385             390             395             400

Gly Phe Gln Gly Gln Gln Pro Pro Leu Ser Gln Val Phe Gln Gly Ile
            405             410             415

Ser Gln Leu Pro Gln Tyr Asn Asn Cys Pro Ser Pro Gln Ala Ala Val
        420             425             430

Gln Gln
```

```
<210>    9
<211>    666
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    9
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1               5               10              15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
            20              25              30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
        35              40              45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
    50              55              60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65              70              75              80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
                85              90              95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
            100             105             110

Ile Asp Cys Asn Glu Asn Thr Gly Lys Lys Ser Gln Lys Glu Thr Glu
        115             120             125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
    130             135             140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145             150             155             160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
            165             170             175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
            180             185             190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
```

```
                195                    200                    205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
    210                 215                 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225                 230                 235                 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
                245                 250                 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
                260                 265                 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
            275                 280                 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
    290                 295                 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305                 310                 315                 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
                325                 330                 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
            340                 345                 350

Gln Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
            355                 360                 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
    370                 375                 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385                 390                 395                 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
                405                 410                 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
            420                 425                 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
    435                 440                 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
    450                 455                 460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465                 470                 475                 480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
                485                 490                 495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
            500                 505                 510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
    515                 520                 525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
```

Page 11

```
      530                    535                    540
Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545                 550                 555                 560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
                565                 570                 575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
            580                 585                 590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
            595                 600                 605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
        610                 615                 620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625                 630                 635                 640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
                645                 650                 655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
            660                 665
```

```
<210>    10
<211>    1000
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    10
atgggcaacc attgtcggga aacgagcaaa ggggccagcc tcaggcccca caacaaactg    60
gggcattccc aattcagcca tttgttcctc agggttttca gggacaacaa cccccactgt    120
cccaagtgtt tcagggaata agccagttac cacaatacaa caattgtccc ccgccacaag    180
cggcagtgca gcagtagatt tatgtactat acaagcagtc tctctgcttc caggggagcc    240
cccacaaaaa accccacag gggtatatgg acccctgcct aaggggactg taggactaat    300
cttgggacga tcaagtctaa atctaaaagg agttcaaatt catactagtg tggttgattc    360
agactataaa ggcgaaattc aattggttat tagctcttca attccttgga gtgccagtcc    420
aagagacagg attgctcaat tattactcct gccatacatt aagggtggaa atagtgaaat    480
aaaaagaata ggagggcttg gaagcactga tccaacagga aaggctgcat attgggcaag    540
tcaggtctca gagaacagac ctgtgtgtaa ggccattatt caaggaaaac agtttgaagg    600
gttggtagac actggagcag atgtctctat cattgcttta aatcagtggc caaaaaattg    660
gcctaaacaa aaggctgtta caggacttgt cggcataggc acagcctcag aagtgtatca    720
aagtacggag attttacatt gcttagggcc agataatcaa gaaagtactg ttcagccaat    780
gattacttca attcctctta atctgtgggg tcgagattta ttacaacaat ggggtgcgga    840
aatcaccatg cccgctccat catatagccc cacgagtcaa aaaatcatga ccaagatggg    900
atatatacca ggaaagggac tagggaaaaa tgaagatggc attaaaattc cagttgaggc    960
taaaataaat caagaaagag aaggaatagg gaatccttgc                          1000
```

```
<210>    11
<211>    1004
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    11
atgggcaacc attgtcggga aacaggaaga ggggccagcc tcaggccccc caacaaactg    60
gggcattccc agttcaactg tttgttcctc agggttttca aggacaacaa cccctacaga    120
aaataccacc acttcaggga gtcagccaat tacaacaatc caacagctgt cccgcgccac    180
agcaggcagc gccacagtag attttatgttc cacccaaatg gtctctttac tccctggaga    240
gcccccacaa aagattccta gaggggtata tggcccgctg ccagaaggga gggtaggcct    300
tattttaggg agatcaagtc taaatttgaa gggagtccaa attcatactg gggtaattta    360
ttcagattat aaagggggaa ttcagttagt gatcagctcc actgttccct ggagtgccaa    420
```

```
tccaggtgat agaattgctc aattactgct tttgccttat gttaaaattg gggaaaacaa    480
aacggaaaga acaggagggt ttggaagtac caaccctgca ggaaaagcca cttattgggc    540
taatcaggtc tcagaggata gacccgtgtg tacagtcact attcagggaa agagtttgaa    600
ggattagtgg atacccaggc tgatgtttct atcatcggca taggcaccgc ctcagaagtg    660
tatcaaagtg ccatgatttt acattgtcta ggatctgata atcaagaaag tacggttcag    720
cctatgatca cttctattcc aatcaattta tggggccgag acttgttaca caatggcat     780
gcagagatta ctatcccagc ctccctatac agccccagga atcaaaaaat catgactaaa    840
atgggatagc tccctaaaaa gggactagga aagaatgaag atggcattaa agtcccaact    900
gaggctgaaa aaaatcaaaa aaagaaaagg aataggggcat ccttttttaga agcggtcact    960
gtagagcctc caaacccat tccattaatt tggggggaaa aaaa                      1004
```

<210> 12
<211> 279
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 12
```
atggagattt tacattgctt agggccagat aatcaagaaa gtactgttca gccaatgatt    60
acttcaattc ctcttaatct gtggggtcga gatttattac aacaatgggg tgcggaaatc    120
accatgcccg ctccattata tagccccacg agtcaaaaaa tcatgaccaa gatgggatat    180
ataccaggaa agggactagg gaaaaatgaa gatggcatta aagttccagt tgaggctaaa    240
ataaatcaag aaagagaagg aataggggtat ccttttttag                         279
```

<210> 13
<211> 92
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 13
```
Met Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr Val
1               5                  10                  15

Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp Leu
            20                  25                  30

Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Leu Tyr Ser
        35                  40                  45

Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly Lys
    50                  55                  60

Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Val Pro Val Glu Ala Lys
65                  70                  75                  80

Ile Asn Gln Glu Arg Glu Gly Ile Gly Tyr Pro Phe
                85                  90
```

<210> 14
<211> 333
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 14
```
Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly Pro
1               5                  10                  15

Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly Phe
            20                  25                  30

Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys Pro
        35                  40                  45

Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala Ala
```
Page 13

```
              50                    55                   60
Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu Pro
65              70              75              80

Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly Thr
                85              90              95

Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val Gln
            100             105             110

Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln Leu
        115             120             125

Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg Ile
    130             135             140

Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu Ile
145             150             155             160

Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala Ala
            165             170             175

Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala Ile
            180             185             190

Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp Val
        195             200             205

Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln Lys
    210             215             220

Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr Gln
225             230             235             240

Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr
            245             250             255

Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp
            260             265             270

Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser Tyr
        275             280             285

Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly
    290             295             300

Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu Ala
305             310             315             320

Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys
            325             330
```

<210> 15
<211> 2896
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 15
```
atggcattaa aattccagtt gaggctaaaa taaatcaaga aagagaagga atagggaatc   60
cttgctaggg gcggccactg tagagcctcc taaacccata ccattaactt ggaaaacaga  120
aaaaccagtg tgggtaaatc agtggccgct accaaaacaa aaactggagg ctttacattt  180
attagcaaat gaacagttag aaaagggtca tattgagcct tcgttctcac cttggaattc  240
```

Page 14

```
tcctgtgttt gtaattcaga agaaatcagg caaatggcgt atgttaactg acttaagggc   300
tgtaaacgcc gtaattcaac ccatggggcc tctccaaccc gggttgccct ctccggccat   360
gatcccaaaa gattggcctt taattataat tgatctaaag gattgctttt ttaccatccc   420
tctggcagag caggattgcg aaaaatttgc ctttactata ccagccataa ataataaaga   480
accagccacc aggtttcagt ggaaagtgtt acctcaggga atgcttaata gtccaactat   540
ttgtcagact tttgtaggtc gagctcttca accagttaga gaaaagtttt cagactgtta   600
tattattcat tgtattgatg atattttatg tgctgcagaa acgaaagata aattaattga   660
ctgttataca tttctgcaag cagaggttgc caatgctgga ctggcaatag catctgataa   720
gatccaaacc tctactcctt ttcattattt agggatgcag atagaaaata gaaaaattaa   780
gccacaaaaa atagaaataa gaaaagacac attaaaaaca ctaaatgatt ttcaaaaatt   840
actaggagat attaattgga ttcggccaac tctaggcatt cctacttatg ccatgtcaaa   900
tttgttctct atcttaagag gagactcaga cttaaatagt aaaagaatgt taaccccaga   960
ggcaacaaaa gaaattaaat tagtggaaga aaaaattcag tcagcgcaaa taaatagaat  1020
agatccctta gccccactcc aacttttgat ttttgccact gcacattctc caacaggcat  1080
cattattcaa aatactgatc ttgtggagtg gtcattcctt cctcacagta cagttaagac  1140
ttttacattg tacttggatc aaatagctac attaatcggt cagacaagat tacgaataat  1200
aaaattatgt gggaatgacc cagacaaaat agttgtccct ttaaccaagg aacaagttag  1260
acaagccttt atcaattctg gtgcatggaa gattggtctt gctaattttg tgggaattat  1320
tgataatcat tacccaaaaa caaagatctt ccagttctta aaattgacta cttggattct  1380
acctaaaatt accagacgtg aacctttaga aaatgctcta acagtattta ctgatggttc  1440
cagcaatgga aaagcagctt acacaggacc gaaagaacga gtaatcaaaa ctccatatca  1500
atcggctcaa agagcagagt tggttgcagt cattacagtg ttacaagatt ttgaccaacc  1560
tatcaatatt atatcagatt ctgcatatgt agtacaggct acaagggatg ttgagacagc  1620
tctaattaaa tatagcatgg atgatcagtt aaaccagcta ttcaatttat tacaacaaac  1680
tgtaagaaaa agaaatttcc cattttatat tacacatatt cgagcacaca ctaatttacc  1740
agggcctttg actaaagcaa atgaacaagc tgacttactg gtatcatctg cactcataaa  1800
agcacaagaa cttcatgctt tgactcatgt aaatgcagca ggattaaaaa acaaatttga  1860
tgtcacatgg aaacaggcaa aagatattgt acaacattgc acccagtgtc aagtcttaca  1920
cctgcccact caagaggcag gagttaatcc cagaggtctg tgtcctaatg cattatggca  1980
aatggatgtc acgcatgtac cttcatttgg aagattatca tatgttcacg taacagttga  2040
tacttattca catttcatat gggcaacttg ccaaacagga gaaagtactt cccatgttaa  2100
aaaacattta ttgtcttgtt ttgctgtaat gggagttcca gaaaaaatca aaactgacaa  2160
tggaccagga tattgtagta aagctttcca aaaattctta agtcagtgga aatttcaca   2220
tacaacagga attccttata attcccaagg acaggccata gttgaaagaa ctaatagaac  2280
actcaaaact caattagtta aacaaaaaga aggggagac agtaaggagt gtaccactcc   2340
tcagatgcaa cttaatctag cactctatac tttaaatttt ttaaacattt atagaaatca  2400
gactactact tctgcagaac aacatcttac tggtaaaaag aacagcccac atgaaggaaa  2460
actaatttgg tggaaagata ataaaaataa gacatgggaa atagggaagg tgataacgtg  2520
ggggagaggt tttgcttgtg tttcaccagg agaaatcag cttcctgttt ggatacccac   2580
tagacatttg aagttctaca atgaacccat cagagatgca aagaaaagca cctccgcgga  2640
gacggagaca tcgcaatcga gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag  2700
aagaacagat gaagttgcca tccaccaaga aggcagagcc gccaacttgg gcacaactaa  2760
agaagctgac gcagttagct acaaaatatc tagagaacac aaaggtgaca caaaccccag  2820
agagtatgct gcttgcagcc ttgatgattg tatcaatggt ggtaagtctc cctatgcctg  2880
caggagcagc tgcagc                                                   2896
```

<210> 16
<211> 2619
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 16

```
atgttaactg acttaagggc tgtaaacgcc gtaattcaac ccatggggcc tctccaaccc    60
gggttgccct ctccggccat gatcccaaaa gattggcctt taattataat tgatctaaag   120
gattgctttt ttaccatccc tctggcagag caggattgcg aaaaatttgc ctttactata   180
ccagccataa ataataaaga accagccacc aggtttcagt ggaaagtgtt acctcaggga   240
atgcttaata gtccaactat ttgtcagact tttgtaggtc gagctcttca accagttaga   300
gaaaagtttt cagactgtta tattattcat tgtattgatg atattttatg tgctgcagaa   360
acgaaagata aattaattga ctgttataca tttctgcaag cagaggttgc caatgctgga   420
ctggcaatag catctgataa gatccaaacc tctactcctt ttcattattt agggatgcag   480
atagaaaata gaaaaattaa gccacaaaaa atagaaataa gaaaagacac attaaaaaca   540
ctaaatgatt ttcaaaaatt actaggagat attaattgga ttcggccaac tctaggcatt   600
cctacttatg ccatgtcaaa tttgttctct atcttaagag gagactcaga cttaaatagt   660
```

```
aaaagaatgt taaccccaga ggcaacaaaa gaaattaaat tagtggaaga aaaaattcag   720
tcagcgcaaa taaatagaat agatcccctta gccccactcc aactttttgat ttttgccact   780
gcacattctc caacaggcat cattattcaa aatactgatc ttgtggagtg gtcattcctt   840
cctcacagta cagttaagac ttttacattg tacttggatc aaatagctac attaatcggt   900
cagacaagat tacgaataat aaaattatgt gggaatgacc cagacaaaat agttgtccct   960
ttaaccaagg aacaagttag acaagccttt atcaattctg gtgcatggaa gattggtctt  1020
gctaattttg tgggaattat tgataatcat tacccaaaaa caaagatctt ccagttctta  1080
aaattgacta cttggattct acctaaaatt accagacgtg aacctttaga aaatgctcta  1140
acagtattta ctgatggttc cagcaatgga aaagcagctt acacaggacc gaaagaacga  1200
gtaatcaaaa ctccatatca atcggctcaa agagcagagt tggttgcagt cattacagtg  1260
ttacaagatt ttgaccaacc tatcaatatt atatcagatt ctgcatatgt agtacaggct  1320
acaagggatg ttgagacagc tctaattaaa tatagcatgg atgatcagtt aaaccagcta  1380
ttcaatttat tacaacaaac tgtaagaaaa agaaatttcc cattttatat tacacatatt  1440
cgagcacaca ctaatttacc agggcctttg actaaagcaa atgaacaagc tgacttactg  1500
gtatcatctg cactcataaa agcacaagaa cttcatgctt tgactcatgt aaatgcagca  1560
ggattaaaaa acaaatttga tgtcacatgg aaacaggcaa aagatattgt acaacattgc  1620
acccagtgtc aagtcttaca cctgcccact caagaggcag gagttaatcc cagaggtctg  1680
tgtcctaatg cattatggca aatggatgtc acgcatgtac cttcatttgg aagattatca  1740
tatgttcacg taacagttga tacttattca catttcatat gggcaacttg ccaaacagga  1800
gaaagtactt cccatgttaa aaaacattta ttgtcttgtt ttgctgtaat gggagttcca  1860
gaaaaaatca aaactgacaa tggaccagga tattgtagta aagctttcca aaaattctta  1920
agtcagtgga aaatttcaca tacaacagga attccttata attcccaagg acaggccata  1980
gttgaaagaa ctaatagaac actcaaaact caattagtta aacaaaaaga aggggggagac  2040
agtaaggagt gtaccactcc tcagatgcaa cttaatctag cactctatac tttaaatttt  2100
ttaaacattt atagaaatca gactactact tctgcagaac aacatcttac tggtaaaaag  2160
aacagcccac atgaaggaaa actaatttgg tggaaagata gtaaaaataa gacatgggaa  2220
ataggggaagg tgataacgtg ggggagaggt tttgcttgtg tttcaccagg agaaaatcag  2280
cttcctgttt ggatacccac tagacatttg aagttctaca atgaacccat cagagatgca  2340
aagaaaagca cctccgcgga gacggagaca tcgcaatcga gcaccgttga ctcacaagat  2400
gaacaaaatg gtgacgtcag aagaacagat gaagttgcca tccaccaaga aggcagagcc  2460
gccaacttgg gcacaactaa agaagctgac gcagttagct acaaaatatc tagagaacac  2520
aaaggtgaca caaaccccag agagtatgct gcttgcagcc ttgatgattg tatcaatggt  2580
ggtaagtctc cctatgcctg caggagcagc tgcagctaa                          2619
```

```
<210>   17
<211>   2671
<212>   DNA
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   17
atggcattaa agtcccaact gaggctgaaa aaaatcaaaa aaagaaaagg aatagggcat    60
ccttttttaga agcggtcact gtagagcctc caaaacccat tccattaatt tgggggggaaa   120
aaaaaaactg tatggtaaat cagtagccgc ttccaaaaca aaaactggag gctttacact   180
tattagcaaa gaaacagtta gaaaaaggac atattgagcc ttcattttcg ccttggaatt   240
ctcctgtttg taattcagaa aaaatccggc agatggcgta tgctaactga cttaagagcc   300
attaatgcca taattcaacc catggggggct ctcccatccc ggttgccctc tccagccatg   360
gtcccctttta attataattg atctgaagga ttgcttttttt accattcctc tggcaaaaga   420
ggattttgaa aaatttgctt ttactatacc agcctaaata ataaagaacc agccaccagg   480
tttcagtgga aagtattgcc tcagggaatg cttaataatt caactatttg tcagactttc   540
atagctcaag ctctgcaacc agttagagac aagttttcag actgttatat cgttcattat   600
gttgatattt tgtgtgctgc agaaacgaga gacaaattaa ttgaccgtta cacatttctc   660
agacagaggt tgccaacgcg ggactgacaa tagcatctga taagattcaa acctctcctc   720
ctttccatta cttggggaatg caggtagagg aaaggaaaat taaaccacaa aaaatagaaa   780
taagaaaaga cacattaaaa acattaaatg agtttcaaaa gttggtagga gatactaatt   840
ggattcggag atattaattg gatttggcca actctaggca ttcctactta tgccatgtca   900
attttgttct ctttcttaag aggggacttg gaattaaata gtgaagaat gttacctcca   960
gaggcaacta aagaaattaa attaattgaa gaaaaaaatt cggtcagcac aagtaaatag  1020
gatcacttgg ccccactcca aattttgatt tttggtactg cacattctct aacagccatc  1080
attgttcaaa acacagatct tgtggattgg tccttccttc ctcatagtac aattaagact  1140
tttacattgt acttggatca aatggctaca ttaattggtc agggaagatt acgaataata  1200
acattgtgtg aaaatgaccc agataaaatc actgttcctt tcaacaagca acaagttaga  1260
caagccttta tcagttctgg tgcatggcag attggtcttg ctaattttct gggaattatt  1320
gataatcatt acccaaaaac aaaaatcttc cagttcttaa aattgactac ttggattcta  1380
```

```
cctaaaatta ccagacgtga accttttagaa aatgctctaa cagtatttac tgatggttcc   1440
agcaatggaa aagcggctta cacagggccg aaagaacgag taatcaaaac tccgtatcaa   1500
tcagctcaaa gagcagagtt ggttgcagtc attacagtgt tacaagattt tgaccaacct   1560
atcaatatta tatcagattc tgcatatgta gtacaggcta caagggatgt tgagacagct   1620
ctaattaaat atagcacgga cgatcattta aaccagctat tcaatttatt acaacaaact   1680
gtaagaaaaa gaaatttccc attttatatt actcatattc gagcacacac taatttacca   1740
gggcctttga ctaaagcaaa tgaacaagct gacttactgg tatcatctgc attcataaaa   1800
gcacaagaac ttcttgcttt gactcatgta aatgcagcag gattaaaaaa caaatttgat   1860
gtcacatgga aacaggcaaa agatattgta caacattgca cccagtgtca agtcttacac   1920
ctgtccactc aagaggcagg agttaatccc agaggtctgt gtcctaatgc gttatggcaa   1980
atggatggca cgcatgttcc ttcatttgga agattatcat atgttcatgt aacagttgat   2040
acttattcac atttcatatg ggcaacttgc caaacaggag aaagtacttc ccatgttaaa   2100
aaacatttat tatcttgttt tgctgtaatg ggagttccag aaaaaatcaa aactgacaat   2160
ggaccaggat attgtagtaa agctttccaa aaattcttaa gtcagtggaa aatttcacat   2220
acaacaggaa ttccttataa ttcccaagga caggccatag ttgaaagaac taatagaaca   2280
ctcaaaactc aattagttaa acaaaaagaa gggggagaca gtaaggagtg taccactcct   2340
cagatgcaac ttaatctagc actctatact ttaaattttt taaacattta tagaaatcag   2400
actactactt ctgcaaaaca acatcttact ggtaaaaagc acagcccaca tgaaggaaaa   2460
ctaatttggt ggaaagataa taaaaataag acatgggaaa tagggaaggt gataacgtgg   2520
gggagaggtt ttgcttgtgt ttcaccagga gaaaatcagc ttcctgtttg gatacccact   2580
agacatttga agttctacaa tgaacccatc ggagatgcaa agaaaagggc ctccacagag   2640
atggtaaccc cagtcacatg gatggataat c                                  2671
```

```
<210>    18
<211>    4086
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    18
atggggcctc tccaacccgg gttgccctct ccggccatga tcccaaaaga ttggccttta    60
attataattg atctaaagga ttgctttttt accatccctc tggcagagca ggattgtgaa   120
aaatttgcct ttactatacc agccataaat aataaagaac cagccaccag gtttcagtgg   180
aaagtgttac ctcagggaat gcttaatagt ccaactattt gtcagacttt tgtaggtcga   240
gctcttcaac cagtgagaga aaagtttca gactgttata ttattcatta tattgatgat   300
atttttatgtg ctgcagaaac gaaagataaa ttaattgact gttatacatt tctgcaagca   360
gaggttgcca atgctggact ggcaatagca tccgataaga tccaaacctc tactccttt    420
cattatttag ggatgcagat agaaaataga aaaattaagc cacaaaaaat agaaataaga   480
aaagacacat taaaaacact aaatgatttt caaaaattac taggagatat taattggatt   540
cggccaactc taggcattc tacttatgcc atgtcaaatt tgttctctat cttaagagga   600
gactcagact taaatagtca aagaatatta accccagagg caacaaaaga aattaaatta.   660
gtggaagaaa aaattcagtc agcgcaaata aatagaatag atcccttagc cccactccaa   720
cttttgattt ttgccactgc acattctcca acaggcatca ttattcaaaa tactgatctt   780
gtggagtggt cattccttcc tcacagtaca gttaagactt ttacattgta cttggatcaa   840
atagctacat taatcggtca gacaagatta cgaataacaa aattatgtgg aaatgaccca   900
gacaaaatag ttgtcccttt aaccaaggaa caagttagac aagcctttat caattctggt   960
gcatggcaga ttggtcttgc taattttgtg ggacttattg ataatcatta cccaaaaaca  1020
aagatcttcc agttcttaaa attgactact tggattctac ctaaaattac cagacgtgaa  1080
cctttagaaa atgctctaac agtatttact gatggttcca gcaatggaaa agcagcttac  1140
acagggccga aagaacgagt aatcaaaact ccatatcaat cggctcaaag agacgagttg  1200
gttgcagtca ttacagtgtt acaagatttt gaccaaccta tcaatattat atcagattct  1260
gcatatgtag tacaggctac aagggatgtt gagacagctc taattaaata tagcatggat  1320
gatcagttaa ccagctatt caatttatta caacaaactg taagaaaaag aaatttccca   1380
ttttatatta cttatattcg agcacacact aatttaccag gcctttgac taaagcaaat  1440
gaacaagctg acttactggt atcatctgca ctcataaag cacaagaact tcatgctttg  1500
actcatgtaa atgcagcagg attaaaaaac aaatttgatg tcacatggaa acaggcaaaa  1560
gatattgtac aacattgcac ccagtgtcaa gtcttacacc tgcccactca gaggcagga   1620
gttaatccca gaggtctgtg tcctaatgca ttatggcaaa tggatgtcac gcatgtacct  1680
tcatttggaa gattatcata tgttcatgta acagttgata cttattcaca tttcatatgg  1740
gcaacttgcc aaacaggaga aagtacttcc catgttaaaa acatttatt gtcttgtttt   1800
gctgtaatgg gagttccaga aaaaatcaaa actgacaatg accaggata ttgtagtaaa   1860
gctttccaaa aattcttaag tcagtggaaa atttcacata acaggaat tccttataat   1920
tcccaaggac aggccatagt tgaaagaact aatagaacac tcaaaactca attagttaaa  1980
caaaaagaag ggggagacag taaggagtgt accactcctc agatgcaact taatctagca  2040
```

```
ctctatactt taaatttttt aaacatttat agaaatcaga ctactacttc tgcagaacaa  2100
catcttactg gtaaaaagaa cagcccacat gaaggaaaac taatttggtg gaaagataat  2160
aaaaataaga catgggaaat agggaaggtg ataacgtggg ggagaggttt tgcttgtgtt  2220
tcaccaggag aaaatcagct tcctgtttgg ttacccacta gacatttgaa gttctacaat  2280
gaacccatcg gagatgcaaa gaaaagggcc tccacggaga tggtaacacc agtcacatgg  2340
atggataatc ctatagaagt atatgttaat gatagtatat gggtacctgg ccccatagat  2400
gatcgctgcc ctgccaaacc tgaggaagaa gggatgatga taaatatttc cattgggtat  2460
cgttatcctc ctatttgcct agggagagca ccaggatgtt taatgcctgc agtccaaaat  2520
tggttggtag aagtacctac tgtcagtccc atcagtagat tcacttatca catggtaagc  2580
gggatgtcac tcaggccacg ggtaaattat ttacaagact tttcttatca aagatcatta  2640
aaatttagac ctaaagggaa accttgcccc aaggaaattc ccaaagaatc aaaaaataca  2700
gaagttttag tttgggaaga atgtgtggcc aatagtgcgg tgatattata aaacaatgaa  2760
tttggaacta ttatagattg ggcacctcga ggtcaattct accacaattg ctcaggacaa  2820
actcagtcgt gtccaagtgc acaagtgagt ccagctgttg atagcgactt aacagaaagt  2880
ttagacaaac ataagcataa aaaattgcag tctttctacc cttgggaatg gggagaaaaa  2940
ggaatctcta ccccaagacc aaaaatagta agtcctgttt ctggtcctga acatccagaa  3000
ttatggaggc ttactgtggc ctcacaccac attagaattt ggtctggaaa tcaaacttta  3060
gaaacaagag attgtaagcc attttatact gtcgacctaa attccagtct aacagttcct  3120
ttacaaagtt gcgtaaagcc cccttatatg ctagttgtag gaaatatagt tattaaacca  3180
gactcccaga ctataacctg tgaaaattgt agattgctta cttgcattga ttcaactttt  3240
aattggcaac accgtattct gctggtgaga gcaagagagg gcgtgtggat ccctgtgtcc  3300
atggaccgac cgtgggaggc ctcaccatcc gtccatattt tgactgaagt attaaaaggt  3360
gttttaaata gatccaaaag attcattttt actttaattg cagtgattat gggattaatt  3420
gcagtcacag ctacggctgc tgtagcagga gttgcattgc actcttctgt tcagtcagta  3480
aactttgtta tgattggca aaagaattct acaagattgt ggaattcaca atctagtatt  3540
gatcaaaaat tggcaaatca aattaatgat cttagacaaa ctgtcatttg gatgggagac  3600
agactcatga gcttagaaca tcgtttccag ttacaatgtg actggaatac gtcagatttt  3660
tgtattacac cccaaatta taatgagtct gagcatcact gggacatggt tagacgccat  3720
ctacagggaa gagaagataa tctcactta gacatttcca aattaaaaga acaaattttc  3780
gaagcatcaa aagcccattt aaatttggtg ccaggaactg aggcaattgc aggagttgct  3840
gatggcctcg caaatcttaa ccctgtcact ggggttaaga ccattggaag tacatcgatt  3900
ataaatctca tattaatcct tgtgtgcctg ttttgtctgt tgttagtctg caggtgtacc  3960
caacagctcc gaagagacag cgaccatcga gaacgggcca tgatgacgat ggcggttttg  4020
tcgaaaagaa aaggggggaaa tgtggggaaa agcaagagag atcaaattgt tactgtgtct  4080
gtgtag                                                             4086
```

```
<210>    19
<211>    872
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    19
Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val Ile Gln Pro Met Gly
1               5                   10                  15


Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys Asp Trp
            20                  25                  30


Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile Pro Leu
        35                  40                  45


Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala Ile Asn
    50                  55                  60


Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro Gln Gly
65                  70                  75                  80


Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg Ala Leu
                85                  90                  95


Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His Cys Ile
            100                 105                 110
```

Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile Asp Cys
        115             120             125

Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala Ile Ala
    130             135             140

Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly Met Gln
145             150             155             160

Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg Lys Asp
            165             170             175

Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp Ile Asn
        180             185             190

Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser Asn Leu
        195             200             205

Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Lys Arg Met Leu
    210             215             220

Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys Ile Gln
225             230             235             240

Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln Leu Leu
            245             250             255

Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln Asn Thr
        260             265             270

Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys Thr Phe
    275             280             285

Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr Arg Leu
    290             295             300

Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val Val Pro
305             310             315             320

Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly Ala Trp
            325             330             335

Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile Asp Asn His Tyr Pro
            340             345             350

Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile Leu Pro
        355             360             365

Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val Phe Thr
    370             375             380

Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys Glu Arg
385             390             395             400

Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala Glu Leu Val Ala
            405             410             415

Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile Ile Ser
            420             425             430

Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr Ala Leu
        435             440             445

Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn Leu Leu
   450              455             460

Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr His Ile
465            470          475            480

Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn Glu Gln
          485          490          495

Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu Leu His
      500          505          510

Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe Asp Val
    515          520          525

Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln Cys Gln
   530          535          540

Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg Gly Leu
545            550          555          560

Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro Ser Phe
          565          570          575

Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser His Phe
        580          585          590

Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val Lys Lys
     595          600          605

His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys Ile Lys
   610          615          620

Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys Phe Leu
625            630          635          640

Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn Ser Gln
          645          650          655

Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr Gln Leu
        660          665          670

Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr Pro Gln
     675          680          685

Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn Ile Tyr
   690          695          700

Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly Lys Lys
705            710          715          720

Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Ser Lys Asn
        725          730          735

Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly Phe Ala
        740          745          750

Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Ile Pro Thr Arg
        755          760          765

His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp Ala Lys Lys Ser Thr
   770          775          780

**Page 20**

Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr Val Asp Ser Gln Asp
785                     790                 795                     800

Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu Val Ala Ile His Gln
                 805                 810                 815

Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys Glu Ala Asp Ala Val
             820                 825                 830

Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp Thr Asn Pro Arg Glu
         835                 840                 845

Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn Gly Gly Lys Ser Pro
    850                 855                 860

Tyr Ala Cys Arg Ser Ser Cys Ser
865                 870

<210> 20
<211> 1361
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<220>
<221> SITE
<222> 917
<223> Xaa is any amino acid

<400> 20
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1               5                   10                  15

Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile
             20                  25                  30

Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala
         35                  40                  45

Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro
    50                  55                  60

Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg
65                  70                  75                  80

Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His
             85                  90                  95

Tyr Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile
             100                 105                 110

Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala
         115                 120                 125

Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly
    130                 135                 140

Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg
145                 150                 155                 160

Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp
             165                 170                 175

Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser
Page 21

```
              180                   185                   190
Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Gln Arg
        195                   200                   205
Ile Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys
    210                   215                   220
Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln
225                   230                   235                   240
Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln
            245                   250                   255
Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys
            260                   265                   270
Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr
        275                   280                   285
Arg Leu Arg Ile Thr Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val
    290                   295                   300
Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly
305                   310                   315                   320
Ala Trp Gln Ile Gly Leu Ala Asn Phe Val Gly Leu Ile Asp Asn His
            325                   330                   335
Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile
            340                   345                   350
Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val
            355                   360                   365
Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys
    370                   375                   380
Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu
385                   390                   395                   400
Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile
            405                   410                   415
Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr
            420                   425                   430
Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn
            435                   440                   445
Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr
    450                   455                   460
Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn
465                   470                   475                   480
Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu
            485                   490                   495
Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe
            500                   505                   510
Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln
```

```
                   515                    520                    525
Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg
    530               535               540
Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro
545               550               555               560
Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser
              565               570               575
His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val
          580               585               590
Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys
          595               600               605
Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys
    610               615               620
Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn
625               630               635               640
Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr
              645               650               655
Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr
              660               665               670
Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn
    675               680               685
Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly
    690               695               700
Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn
705               710               715               720
Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly
              725               730               735
Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro
              740               745               750
Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys
              755               760               765
Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro
    770               775               780
Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp
785               790               795               800
Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile
              805               810               815
Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly
          820               825               830
Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val
    835               840               845
Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu
```

```
            850                    855                    860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865                    870                875                    880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
                885                890                895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
                900                905                910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
                915                920                925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
    930                935                940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945                950                955                    960

Leu Asp Lys His Lys His Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu
                965                970                975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
                980                985                990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
        995                1000               1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
    1010               1015               1020

Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
1025               1030               1035               1040

Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
        1045               1050               1055

Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
        1060               1065               1070

Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
        1075               1080               1085

Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
    1090               1095               1100

Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
1105               1110               1115               1120

Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
                1125               1130               1135

Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala Val Ala Gly Val Ala
        1140               1145               1150

Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
        1155               1160               1165

Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
    1170               1175               1180

Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp
```

Page 24

```
        1185              1190              1195              1200

Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
                 1205              1210              1215

Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
             1220              1225              1230

His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
         1235              1240              1245

Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
     1250              1255              1260

Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
1265              1270              1275              1280

Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
             1285              1290              1295

Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
             1300              1305              1310

Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
         1315              1320              1325

His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
     1330              1335              1340

Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
1345              1350              1355              1360

Val
```

<210> 21
<211> 956
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 21

```
Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala Ala
1               5                  10                 15

Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu Lys
             20                 25                 30

Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu Ala
         35                 40                 45

Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu Pro
     50                 55                 60

Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys Ser
65                 70                 75                 80

Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val Ile
             85                 90                 95

Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile
             100                105                110

Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe
```

```
              115                    120                    125
Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile
    130                    135                    140
Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val
145             150                    155                    160
Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val
                165                    170                    175
Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile
                180                    185                    190
Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys
            195                    200                    205
Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly
    210                    215                    220
Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr
225             230                    235                    240
Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu
                245                    250                    255
Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu
            260                    265                    270
Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala
        275                    280                    285
Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser
    290                    295                    300
Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu
305                 310                    315                    320
Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro
                325                    330                    335
Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile
            340                    345                    350
Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr
        355                    360                    365
Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly
    370                    375                    380
Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp Lys
385                 390                    395                    400
Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn
            405                    410                    415
Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile Asp
        420                    425                    430
Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr
        435                    440                    445
Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu
```

```
                    450                      455                     460

Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly
465                 470                 475                     480

Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala
                485                 490                 495

Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile
            500                 505                 510

Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val
            515                 520                 525

Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu
    530                 535                 540

Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr
545                 550                 555                     560

Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys
                565                 570                 575

Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala
            580                 585                 590

Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn
    595                 600                 605

Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys
610                 615                 620

Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn
625                 630                 635                     640

Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His
                645                 650                 655

Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr
                660                 665                 670

Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser
    675                 680                 685

His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro
    690                 695                 700

Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe
705                 710                 715                     720

Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro
                725                 730                 735

Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu
                740                 745                 750

Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys
    755                 760                 765

Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe
    770                 775                 780

Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu
```

Page 27

```
785                   790                   795                   800
```

Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys
          805                   810              815

Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly
          820                   825                   830

Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp
          835                   840                   845

Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp Ala
      850                   855                   860

Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr Val
865                   870                   875                   880

Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu Val
                  885                   890                   895

Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys Glu
              900                   905                   910

Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp Thr
          915                   920                   925

Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn Gly
      930                   935                   940

Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser
945                   950                   955

```
<210>    22
<211>    2000
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    22
atgaacccat cagagatgca aagaaaagca cctccgcgga gacggagaca tcgcaatcga   60
gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag aagaacagat gaagttgcca  120
tccaccaaga aggcagagcc gccaacttgg gcacaactaa agaagctgac gcagttagct  180
acaaaatatc tagagaacac aaaggtgaca caaaccccag agagtatgct gcttgcagcc  240
ttgatgattg tatcaatggt ggtaagtctc cctatgcctg caggagcagc tgcagctaac  300
tatacctact gggcctatgt gcctttcccg cccttaattc gggcagtcac atggatggat  360
aatcctacag aagtatatgt taatgatagt gtatgggtac ctggccccat agatgatcgc  420
tgccctgcca aacctgagga agaagggatg atgataaata tttccattgg gtatcattat  480
cctcctattt gcctagggag agcaccagga tgtttaatgc ctgcagtcca aaattggttg  540
gtagaagtac ctactgtcag tcccatctgt agattcactt atcacatggt aagcgggatg  600
tcactcaggc acgggtaaa ttatttacaa gacttttctt atcaaagatc attaaaattt  660
agacctaaag ggaaaccttg ccccaaggaa attcccaaag aatcaaaaaa tacagaagtt  720
ttagtttggg aagaatgtgt ggccaatagt gcggtgatat acaaaacaa tgaattcgga  780
actattatag attgggcacc tcgaggtcaa ttctaccaca attgctcagg acaaactcag  840
tcgtgtccaa gtgcacaagt gagtccagct gttgatagcg acttaacaga aagtttagac  900
aaacataagc ataaaaaatt gcagtctttc tacccttggg aatggggaga aaaaggaatc  960
tctaccccaa gaccaaaaat agtaagtcct gtttctggtc ctgaacatcc agaattatgg 1020
aggcttactg tggcctcaca ccacattaga atttggtctg gaaatcaaac tttagaaaca 1080
agagatcgta agccatttta tactattgac ctgaattcca gtctaacagt tcctttacaa 1140
agttgcgtaa agccccctta tatgctagtt gtaggaaata tagttattaa accagactcc 1200
cagactataa cctgtgaaaa ttgtagattg cttacttgca ttgattcaac ttttaattgg 1260
caacaccgta ttctgctggt gagagcaaga gagggcgtgt ggatccctgt gtccatggac 1320
cgaccgtggg aggcctcgcc atccgtccat attttgactg aagtattaaa aggtgtttta 1380
aatagatcca aaagattcat ttttactta attgcagtga ttatgggatt aattgcagtc 1440
acagctacgg ctgctgtagc aggagttgca ttgcactctt ctgttcagtc agtaaacttt 1500
```

```
gttaatgatt ggcaaaaaaa ttctacaaga ttgtggaatt cacaatctag tattgatcaa    1560
aaattggcaa atcaaattaa tgatcttaga caaactgtca tttggatggg agacagactc    1620
atgagcttag aacatcgttt ccagttacaa tgtgactgga atacgtcaga tttttgtatt    1680
acaccccaaa tttataatga gtctgagcat cactgggaca tggttagacg ccatctacag    1740
ggaagagaag ataatctcac tttagacatt tccaaattaa aagaacaaat tttcgaagca    1800
tcaaaagccc atttaaattt ggtgccagga actgaggcaa ttgcaggagt tgctgatggc    1860
ctcgcaaatc ttaaccctgt cacttgggtt aagaccattg gaagtactac gattataaat    1920
ctcatattaa tccttgtgtg cctgttttgt ctgttgttag tctgcaggtg tacccaacag    1980
ctccgaagag acagcgacca                                                 2000
```

```
<210>    23
<211>    2085
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    23
atgcaaagaa aagcacctcc gcggagacgg agacatcgca atcgagcacc gttgactcac    60
aagatgaaca aaatggtgac gtcagaagaa cagatgaagt tgccatccac caagaaggca    120
gagccgccaa cttgggcaca actaaagaag ctgacgcagt tagctacaaa atatctagag    180
aacacaaagg tgacacaaac cccagagagt atgctgcttg cagccttgat gattgtatca    240
atggtggtaa gtctccctat gcctgcagga gcagctgcag ctaactatac ctactgggcc    300
tatgtgcctt tcccgccctt aattcgggca gtcacatgga tggataatcc tacagaagta    360
tatgttaatg atagtgtatg ggtacctggc cccatagatg atcgctgccc tgccaaacct    420
gaggaagaag ggatgatgat aaatatttcc attgggtatc attatcctcc tatttgccta    480
gggagagcac caggatgttt aatgcctgca gtccaaaatt ggttggtaga agtacctact    540
gtcagtccca tctgtagatt cacttatcac atggtaagcg ggatgtcact caggccacgg    600
gtaaattatt tacaagactt ttcttatcaa agatcattaa aatttagacc taaagggaaa    660
ccttgcccca aggaaattcc caaagaatca aaaaatacag aagtttttagt ttgggaagaa    720
tgtgtggcca atagtgcggt gatattacaa aacaatgaat tcggaactat tatagattgg    780
gcacctcgag gtcaattcta ccacaattgc tcaggacaaa ctcagtcgtg tcaaagtgca    840
caagtgagtc cagctgttga tagcgactta acagaaagtt tagacaaaca taagcataaa    900
aaattgcagt ctttctaccc ttgggaatgg ggagaaaaag gaatctctac cccaagacca    960
aaaatagtaa gtcctgtttc tggtcctgaa catccagaat tatggaggct tactgtggcc    1020
tcacaccaca ttagaatttg gtctggaaat caaactttag aaacaagaga tcgtaagcca    1080
ttttatacta ttgacctgaa ttccagtcta acagttcctt tacaaagttg cgtaaagccc    1140
ccttatatgc tagttgtagg aaatatagtt attaaaccag actcccagac tataacctgt    1200
gaaaattgta gattgcttac ttgcattgat tcaacttttta attggcaaca ccgtattctg    1260
ctggtgagag caagagaggg cgtgtggatc cctgtgtcca tggaccgacc gtgggaggcc    1320
tcgccatccg tccatatttt gactgaagta ttaaaaggtg ttttaaatag atccaaaaga    1380
ttcatttttta ctttaattgc agtgattatg ggattaattg cagtcacagc tacggctgct    1440
gtagcaggag ttgcattgca ctcttctgtt cagtcagtaa actttgttaa tgattggcaa    1500
aaaaattcta caagatgtgt gaattcacaa tctagtattg atcaaaatcaa             1560
attaatgatc ttagacaaac tgtcatttgg atgggagaca gactcatgag cttagaacat    1620
cgtttccagt tacaatgtga ctggaatacg tcagattttt gtattacacc ccaaatttat    1680
aatgagtctg agcatcactg ggacatggtt agacgccatc tacagggaag agaagataat    1740
ctcactttag acatttccaa attaaaagaa caaattttcg aagcatcaaa agcccattta    1800
aatttggtgc caggaactga ggcaattgca ggagttgctg atggcctcgc aaatcttaac    1860
cctgtcactt gggttaagac cattggaagt actacgatta taaatctcat attaatcctt    1920
gtgtgcctgt tttgtctgtt gttagtctgc aggtgtaccc aacagctccg aagagacagc    1980
gaccatcgag aacgggccat gatgacgatg gcggttttgt cgaaaagaaa aggggggaaat    2040
gtggggaaaa gcaagagaga tcagattgtt actgtgtctg tgtag                    2085
```

```
<210>    24
<211>    1665
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    24
gtcacatgga tggataatcc tatagaagta tatgttaatg atagtgtatg ggtacctggc    60
cccacagatg atcgctgccc tgccaaacct gaggaagaag ggatgatgat aaatatttcc    120
attgtgtatc gttatcctcc tatttgccta gggagagcac caggatgttt aatgcctgca    180
gtccaaaatt ggttggtaga agtacctact gtcagtccta acagtagatt cacttatcac    240
atggtaagcg ggatgtcact caggccacgg gtaaattatt tacaagactt ttcttatcaa    300
```

Page 29

```
agatcattaa aatttagacc taaagggaaa ccttgcccca aggaaattcc caaagaatca    360
aaaaatacag aagtttttagt ttgggaagaa tgtgtggcca atagtgcggt gatattacaa    420
aacaatgaat tcggaactat tatagattgg gcacctcgag gtcaattcta ccacaattgc    480
tcaggacaaa ctcagtcgtg tccaagtgca caagtgagtc cagctgttga tagcgactta    540
acagaaagtc tagacaaaca taagcataaa aaattacagt ctttctaccc ttgggaatgg    600
ggagaaaaag gaatctctac cccaagacca gaaataataa gtcctgtttc tggtcctgaa    660
catccagaat tatggaggct ttggcctgac accacattag aatttggtct ggaaatcaaa    720
ctttagaaac aagagatcgt aagccatttt atactatcga cctaaattcc agtctaacgg    780
ttcctttaca aagttgcgta aagccctctt atatgctagt tgtaggaaat atagttatta    840
aaccagactc ccaaactata acctgtgaaa attgtagatt gtttacttgc attgattcaa    900
cttttaattg gcggcaccgt attctgctgg tgagagcaag agagggcgtg tggatctctg    960
tgtccgtgga ctgaccgtgg gaggcctcgc catccatcca tattttgact gaagtattaa   1020
aagacatttt aaatagatcc aaaagattca tttttacctt aattgcagtg attatgggat   1080
taattgcagt cacagctacg gctgctgtgg caggagttgc attgcactct tctgttcagt   1140
cggtaaactt tgttaatgat tggcaaaaga attctacaag attgtggaat tcacaatcta   1200
gtattgatca aaaattggca aatcaaatta atgatcttag acaaactgtc atttggatgg   1260
gagacagact catgagctta gaacattgtt ccagttaca gtgtgactgg aatacgtcag   1320
attttttgtat tacacccaa atttataatg agtctgagca tcactgggac atggttagac   1380
gccatctaca gggaagagaa gataatctca ctttagacat ttccaaatta aaataacaaa   1440
ttttcgaagc atcaaaagcc catttaaatt tgatgccagg aactgaggca attgcaggag   1500
ttgctgatgg cctcgcaaat cttaaccctg tcacttgggt taagaccatc ggaagtacta   1560
tgattataaa tctcatatta atccttgtgt gcctgttttg tctgttgtta gtctgcaggt   1620
gtacccaaca gctccgaaga gacagcgacc atcgagaacg ggcca               1665
```

```
<210>    25
<211>    4086
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    25
atggggcctc tccaacccgg gttgccctct ccggccatga tcccaaaaga ttggccttta     60
attataattg atctaaagga ttgcttttttt accatccctc tggcagagca ggattgtgaa    120
aaatttgcct ttactatacc agccataaat aataaagaac cagccaccag gtttcagtgg    180
aaagtgttac ctcagggaat gcttaatagt ccaactattt gtcagacttt tgtaggtcga    240
gctcttcaac cagtgagaga aaagtttttca gactgttata ttattcatta tattgatgat    300
atttttatgtg ctgcagaaac gaaagataaa ttaattgact gttatacatt tctgcaagca    360
gaggttgcca atgctggact ggcaatagca tccgataaga tccaaacctc tactcctttt    420
cattatttag ggatgcagat agaaaataga aaaattaagc cacaaaaaat agaaataaga    480
aaagacacat taaaaacact aaatgatttt caaaaattac taggagatat taattggatt    540
cggccaactc taggcattcc tacttatgcc atgtcaaatt tgttctctat cttaagagga    600
gactcagact taaatagtca aagaatatca accccagagg caacaaaaga aattaaatta    660
gtggaagaaa aaattcagtc agcgcaaata aatagaatag atcccttagc cccactccaa    720
ctttttgattt ttgccactgc acattctcca acaggcatca ttattcaaaa tactgatctt    780
gtggagtggt cattccttcc tcacagtaca gttaagactt ttacattgta cttggatcaa    840
atagctacat taatcggtca gacaagatta cgaataacaa aattatgtgg aaatgaccca    900
gacaaaatag ttgtcccttt aaccaggaa caagttagac aagcctttat caattctggt    960
gcatggcaga ttggtcttgc taattttgtg ggacttattg ataatcatta cccaaaaaca   1020
aagatcttcc agttcttaaa attgactact tggattctac ctaaaattac cagacgtgaa   1080
cctttagaaa atgctctaac agtatttact gatggttcca gcaatggaaa agcagcttac   1140
acagggccga aagaacgagt aatcaaaact ccatatcaat cggctcaaag agacgagttg   1200
gttgcagtca ttacagtgtt acaagatttt gaccaaccta tcaatattat atcagattct   1260
gcatatgtag tacaggctac aagggatgtt gagacagctc taattaaata tagcatggat   1320
gatcagttaa accagctatt caatttatta caacaaactg taagaaaaag aaatttccca   1380
ttttatatta cttatattcg agcacacact aatttaccag ggcctttgac taaagcaaat   1440
gaacaagctg acttactggt atcatctgca ctcataaaag cacaagaact tcatgctttg   1500
actcatgtaa atgcagcagg attaaaaaac aaatttgatg tcacatggaa acaggcaaaa   1560
gatattgtac aacattgcac ccagtgtcaa gtcttacacc tgcccactca agaggcagga   1620
gttaatccca gaggtctgtg tcctaatgca ttatggcaaa tggatgtcac gcatgtacct   1680
tcatttggaa gattatcata tgttcatgta acagttgata cttattcaca tttcatatgg   1740
gcaacttgcc aaacaggaga aagtacttcc catgttaaaa aacatttatt gtcttgtttt   1800
gctgtaatgg gagttccaga aaaaatcaaa actgacaatg accaggata ttgtagtaaa   1860
gctttccaaa aattcttaag tcagtggaaa atttcacata aacaggaat tccttataat   1920
tcccaaggac aggccatagt tgaaagaact aatagaacac tcaaaactca attagttaaa   1980
```

```
caaaaagaag ggggagacag taaggagtgt accactcctc agatgcaact taatctagca    2040
ctctatactt taaatttttt aaacatttat agaaatcaga ctactacttc tgcagaacaa    2100
catcttactg gtaaaaagaa cagcccacat gaaggaaaac taatttggtg gaaagataat    2160
aaaaataaga catgggaaat agggaaggtg ataacgtggg ggagaggttt tgcttgtgtt    2220
tcaccaggag aaaatcagct tcctgtttgg ttacccacta gacatttgaa gttctacaat    2280
gaacccatcg gagatgcaaa gaaaagggcc tccacggaga tggtaacacc agtcacatgg    2340
atggataatc ctatagaagt atatgttaat gatagtatat gggtacctgg ccccatagat    2400
gatcgctgcc ctgccaaacc tgaggaagaa gggatgatga taaatatttc cattgggtat    2460
cgttatcctc ctatttgcct agggagagca ccaggatgtt taatgcctgc agtccaaaat    2520
tggttggtag aagtacctac tgtcagtccc atcagtagat tcacttatca catggtaagc    2580
gggatgtcac tcaggccacg ggtaaattat ttacaagact tttcttatca aagatcatta    2640
aaatttagac ctaaagggaa accttgcccc aaggaaattc ccaaagaatc aaaaaataca    2700
gaagttttag tttgggaaga atgtgtggcc aatagtgcgg tgatattata aaacaatgaa    2760
tttggaacta ttatagattg ggcacctcga ggtcaattct accacaattg ctcaggacaa    2820
actcagtcgt gtccaagtgc acaagtgagt ccagctgttg atagcgactt aacagaaagt    2880
ttagacaaac ataagcataa aaaattgcag tctttctacc cttgggaatg gggagaaaaa    2940
ggaatctcta ccccaagacc aaaaatagta agtcctgttt ctggtcctga acatccagaa    3000
ttatggaggc ttactgtggc ctcacaccac attagaattt ggtctggaaa tcaaacttta    3060
gaaacaagag attgtaagcc attttatact gtcgacctaa attccagtct aacagttcct    3120
ttacaaagtt gcgtaaagc cccttatatg ctagttgtag gaaatatagt tattaaacca    3180
gactcccaga ctataacctg tgaaaattgt agattgctta cttgcattga ttcaactttt    3240
aattggcaac accgtattct gctggtgaga gcaagagagg gcgtgtggat ccctgtgtcc    3300
atggaccgac cgtgggaggc ctcaccatcc gtccatattt tgactgaagt attaaaaggt    3360
gtttaaaata gatccaaaag attcattttt actttaattg cagtgattat gggattaatt    3420
gcagtcacag ctacggctgc tgtagcagga gttgcattgc actcttctgt tcagtcagta    3480
aactttgtta atgattggca aaagaattct acaagattgt ggaattcaca atctagtatt    3540
gatcaaaaat tggcaaatca aattaatgat cttagacaaa ctgtcatttg atgggagac     3600
agactcatga gcttagaaca tcgtttccag ttacaatgtg actggaatac gtcagatttt    3660
tgtattacac cccaaattta taatgagtct gagcatcact gggacatggt tagacgccat    3720
ctacagggaa gagaagataa tctcacttta gacatttcca aattaaaaga acaaattttc    3780
gaagcatcaa aagcccattt aaatttggtg ccaggaactg aggcaattgc aggagttgct    3840
gatggcctcg caaatcttaa ccctgtcact tgggttaaga ccattggaag tacatcgatt    3900
ataaatctca tattaatcct tgtgtgcctg ttttgtctgt tgttagtctg caggtgtacc    3960
caacagctcc gaagagacag cgaccatcga gaacgggcca tgatgacgat ggcggttttg    4020
tcgaaaagaa aagggggaaa tgtggggaaa agcaagagag atcaaattgt tactgtgtct    4080
gtgtag                                                               4086
```

<210>  26
<211>  694
<212>  PRT
<213>  Human endogenous retrovirus, K family (HERV-K)

<400>  26
```
Met Gln Arg Lys Ala Pro Pro Arg Arg Arg His Arg Asn Arg Ala
1               5                   10                  15

Pro Leu Thr His Lys Met Asn Lys Met Val Thr Ser Glu Glu Gln Met
            20                  25                  30

Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro Thr Trp Ala Gln Leu
        35                  40                  45

Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu Glu Asn Thr Lys Val
    50                  55                  60

Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala Leu Met Ile Val Ser
65                  70                  75                  80

Met Val Val Ser Leu Pro Met Pro Ala Gly Ala Ala Ala Ala Asn Tyr
                85                  90                  95

Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu Ile Arg Ala Val Thr
            100                 105                 110
```
Page 31

Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn Asp Ser Val Trp Val
        115             120                 125

Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly
    130             135             140

Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr Pro Pro Ile Cys Leu
145             150             155             160

Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val
                165             170                 175

Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe Thr Tyr His Met Val
            180             185             190

Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser
        195             200             205

Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys
    210             215             220

Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu
225             230             235             240

Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn Asn Glu Phe Gly Thr
                245             250             255

Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly
            260             265             270

Gln Thr Gln Ser Cys Gln Ser Ala Gln. Val Ser Pro Ala Val Asp Ser
        275             280             285

Asp Leu Thr Glu Ser Leu Asp Lys His Lys His Lys Lys Leu Gln Ser
290             295             300

Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro
305             310             315             320

Lys Ile Val Ser Pro Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg
            325             330             335

Leu Thr Val Ala Ser His His Ile Arg Ile Trp Ser Gly Asn Gln Thr
            340             345             350

Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr Ile Asp Leu Asn Ser
        355             360             365

Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu
370             375             380

Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys
385             390             395             400

Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln
            405             410             415

His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val
        420             425             430

Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr
        435             440             445

Page 32

```
Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr
    450             455             460

Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala
465             470             475             480

Val Ala Gly Val Ala Leu His Ser Ser Val Gln Ser Val Asn Phe Val
            485             490             495

Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser
        500   ·         505             510

Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val
        515             520             525

Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu
  · 530             535             540

Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr
545             550             555             560

Asn Glu Ser Glu His His Trp Asp Met Val Arg Arg His Leu Gln Gly
            565             570             575

Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile
        580             585             590

Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala
        595             600             605

Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp
    610             615             620

Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn Leu Ile Leu Ile Leu
625             630             635   .         640

Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu
            645             650             655

Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met Met Thr Met Ala Val
        660             665             670   ·

Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln
    675             680             685

Ile Val Thr Val Ser Val
    690
```

```
<210>   27
<211>   1361
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<220>
<221>   SITE
<222>   917
<223>   Xaa is any amino acid

<400>   27
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1               5               10              15
```

Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile
            20              25              30

Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala
            35              40              45

Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro
    50              55              60

Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg
65              70              75              80

Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His
            85              90              95

Tyr Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile
            100             105             110

Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala
        115             120             125

Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly
    130             135             140

Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg
145             150             155             160

Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp
            165             170             175

Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser
        180             185             190

Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Gln Arg
        195             200             205

Ile Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys
    210             215             220

Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln
225             230             235             240

Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln
            245             250             255

Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys
        260             265             270

Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr
        275             280             285

Arg Leu Arg Ile Thr Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val
    290             295             300

Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly
305             310             315             320

Ala Trp Gln Ile Gly Leu Ala Asn Phe Val Gly Leu Ile Asp Asn His
            325             330             335

Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile
        340             345             350

**Page 34**

Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val
        355                 360                 365

Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys
    370                 375                 380

Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu
385                 390                 395                 400

Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile
                405                 410                 415

Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr
            420                 425                 430

Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn
            435                 440                 445

Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr
    450                 455                 460

Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn
465                 470                 475                 480

Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu
            485                 490                 495

Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe
            500                 505                 510

Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln
        515                 520                 525

Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg
    530                 535                 540

Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro
545                 550                 555                 560

Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser
            565                 570                 575

His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val
            580                 585                 590

Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys
        595                 600                 605

Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys
    610                 615                 620

Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn
625                 630                 635                 640

Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr
            645                 650                 655

Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr
            660                 665                 670

Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn
            675                 680                 685

Page 35

Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly
    690             695         700

Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn
705         710             715             720

Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly
            725             730             735

Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro
            740             745             750

Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys
        755             760             765

Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro
    770             775             780

Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp
785             790             795             800

Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile
            805             810             815

Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly
            820             825             830

Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val
        835             840             845

Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu
    850             855             860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865             870             875             880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
            885             890             895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
        900             905             910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
        915             920             925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
    930             935             940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945             950             955             960

Leu Asp Lys His Lys His Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu
            965             970             975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
            980             985             990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
        995             1000            1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
    1010            1015            1020

Page 36

```
Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
1025                1030                1035                1040

Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
                1045                1050                1055

Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
            1060                1065                1070

Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
            1075                1080                1085

Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
        1090                1095                1100

Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
1105                1110                1115                1120

Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
                1125                1130                1135

Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala Val Ala Gly Val Ala
            1140                1145                1150

Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
            1155                1160                1165

Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
        1170                1175                1180

Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp
1185                1190                1195                1200

Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
                1205                1210                1215

Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
            1220                1225                1230

His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
            1235                1240                1245

Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
    1250                1255                1260

Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
1265                1270                1275                1280

Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
            1285                1290                1295

Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
            1300                1305                1310

Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
        1315                1320                1325

His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
    1330                1335                1340

Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
1345                1350                1355                1360
```

**Page 37**

Val

<210> 28
<211> 699
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 28

```
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                  10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
                85                  90                  95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
            100                 105                 110

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
            115                 120                 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
    130                 135                 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145                 150                 155                 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
            165                 170                 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
            180                 185                 190

Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
            195                 200                 205

Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
    210                 215                 220

Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
225                 230                 235                 240

Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
            245                 250                 255

Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
            260                 265                 270

His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
    275                 280                 285
```

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
290                     295             300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305             310             315             320

Ser Thr Pro Arg Pro Lys Ile Val Ser Pro Val Ser Gly Pro Glu His
              325             330             335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
          340             345             350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
          355             360             365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
    370             375             380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385             390             395             400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
              405             410             415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
          420             425             430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
          435             440             445

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
    450             455             460

Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
465             470             475             480

Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
          485             490             495

Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
          500             505             510

Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
          515             520             525

Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
    530             535             540

His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
545             550             555             560

Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
              565             570             575

Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
          580             585             590

Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
          595             600             605

Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
    610             615             620

Page 39

```
Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
625             630             635             640

Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
            645             650             655

Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
        660             665             670

Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
        675             680             685

Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val
    690             695
```

<210> 29
<211> 294
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 29
```
agttctacaa tgaacccatc agagatgcaa agaaaagcac ctccgcggag acggagacat    60
cgcaatcgag caccgttgac tcacaagatg aacaaaatgg tgacgtcaga agaacagatg   120
aagttgccat ccaccaagaa ggcagagccg ccaacttggg cacaactaaa gaagctgacg   180
cagttagcta caaaatatct agagaacaca aaggtgacac aaaccccaga gagtatgctg   240
cttgcagcct tgatgattgt atcaatggtg gtaagtctcc ctatgcctgc agga          294
```

<210> 30
<211> 57
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 30
```
tctgcaggtg tacccaacag ctccgaagag acagcgacca tcgagaacgg gccatga       57
```

<210> 31
<211> 105
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 31
```
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5               10              15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20              25              30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Gly Pro Pro
        35              40              45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50              55              60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65              70              75              80

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
            85              90              95

Glu Thr Ala Thr Ile Glu Asn Gly Pro
            100             105
```

<210> 32

```
<211>    86
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)
```

```
<400>    32
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
65                  70                  75                  80

Thr Ile Glu Asn Gly Pro
                85
```

```
<210>    33
<211>    74
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)
```

```
<400>    33
Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1               5                   10                  15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
            20                  25                  30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
        35                  40                  45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
    50                  55                  60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65                  70
```

```
<210>    34
<211>    79
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)
```

```
<400>    34
Met Asn Ser Leu Glu Met Gln Arg Lys Val Trp Arg Trp Arg His Pro
1               5                   10                  15

Asn Arg Leu Ala Ser Leu Gln Val Tyr Pro Ala Ala Pro Lys Arg Gln
            20                  25                  30

Gln Pro Ala Arg Met Gly His Ser Asp Asp Gly Gly Phe Val Lys Lys
        35                  40                  45

Lys Arg Gly Gly Tyr Val Arg Lys Arg Glu Ile Arg Leu Ser Leu Cys
    50                  55                  60

Leu Cys Arg Lys Gly Arg His Lys Lys Leu His Phe Val Leu Tyr
```

<210>    35
<211>    129
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    35
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Ile Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65                  70                  75                  80

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
            85                  90                  95

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
            100                 105                 110

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu
            115                 120                 125

Tyr


<210>    36
<211>    125
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    36
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65                  70                  75                  80

Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
            85                  90                  95

Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
            100                 105                 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
                                    Page 42

```
<210>   37
<211>   144
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   37
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
                85                  90                  95

Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys
            100                 105                 110

Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
            115                 120                 125

Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
    130                 135                 140


<210>   38
<211>   74
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   38
Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1               5                   10                  15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
            20                  25                  30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
            35                  40                  45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
    50                  55                  60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65                  70


<210>   39
<211>   74
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   39
```

Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1               5               10              15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
            20              25              30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
            35              40              45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
    50              55              60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65                      70

<210>    40
<211>    44
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    40
Met Glu Tyr Lys Asn Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly
1               5               10              15

Asp Ala Lys Lys Arg Ala Ser Thr Glu Met Ser Ala Gly Val Pro Asn
            20              25              30

Ser Ser Glu Glu Thr Ala Thr Ile Glu Asn Gly Pro
            35              40

<210>    41
<211>    74
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    41
Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1               5               10              15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
            20              25              30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
            35              40              45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
    50              55              60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65                      70

<210>    42
<211>    86
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    42
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5               10              15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20              25              30

Page 44

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
65                  70                  75                  80

Thr Ile Glu Asn Gly Pro
                85

<210>   43
<211>   105
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   43
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
                85                  90                  95

Glu Thr Ala Thr Ile Glu Asn Gly Pro
            100                 105

<210>   44
<211>   127
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   44
Met Val Thr Pro Val Thr Trp Met Asp Asn Pro Ile Glu Val Tyr Val
1               5                   10                  15

Asn Asp Ser Val Trp Val Pro Gly Pro Thr Asp Asp Arg Cys Pro Ala
            20                  25                  30

Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Val Tyr Arg
        35                  40                  45

Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala
    50                  55                  60

Val Gln Asn Cys Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg
65                  70                  75                  80

Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys
                85                  90                  95

Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val
        100             105             110

Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
        115             120             125

<210>    45
<211>    105
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    45
Met Val Thr Pro Val Thr Trp Met Asp Asn Pro Ile Glu Val Tyr Val
1                5               10              15

Asn Asp Ser Glu Trp Val Pro Gly Pro Thr Asp Asp Arg Cys Pro Ala
            20              25              30

Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Leu Gln
            35              40              45

Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly His
    50              55              60

Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly Glu
65              70              75              80

Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser Arg
            85              90              95

His Arg Arg Leu His Phe Val Met Cys
            100             105

<210>    46
<211>    79
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    46
Met Asn Ser Leu Glu Met Gln Arg Lys Val Trp Arg Trp Arg His Pro
1                5               10              15

Asn Arg Leu Ala Ser Leu Gln Val Tyr Pro Ala Ala Pro Lys Arg Gln
            20              25              30

Gln Pro Ala Arg Met Gly His Ser Asp Asp Gly Gly Phe Val Lys Lys
            35              40              45

Lys Arg Gly Gly Tyr Val Arg Lys Arg Glu Ile Arg Leu Ser Leu Cys
    50              55              60

Leu Cys Arg Lys Gly Arg His Lys Lys Leu His Phe Asp Leu Tyr
65              70              75

<210>    47
<211>    214
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    47
Met Asn Ser Leu Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1                5               10              15

Page 46

```
His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20              25              30

Ser Glu Glu Gln Met Lys Leu Ser Ser Thr Lys Lys Ala Glu Pro Pro
            35              40              45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
     50              55              60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65              70              75              80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
                85              90              95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
            100             105             110

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
            115             120             125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
    130             135             140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145             150             155             160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
            165             170             175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
            180             185             190

Thr Tyr His Met Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
            195             200             205

Thr Ile Glu Asn Gly Pro
    210
```

```
<210>    48
<211>    129
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    48
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1                5               10              15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20              25              30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
            35              40              45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
     50              55              60

Glu Asn Thr Lys Val Thr Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65              70              75              80

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
            85              90              95
```

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
            100                 105                 110

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met
        115                 120                 125

Tyr


<210>    49
<211>    125
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    49
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65                  70                  75                  80

Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
            85                  90                  95

Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
            100                 105                 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met Tyr
        115                 120                 125

<210>    50
<211>    145
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<220>
<221>    SITE
<222>    64
<223>    Xaa is any amino acid

<400>    50
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Xaa
    50                  55                  60

Leu Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala

Page 48

```
65                      70                   75                    80
Ala Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg
                85                   90                   95

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
            100                  105                  110

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
            115                  120                  125

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met
    130                  135                  140

Tyr
145
```

```
<210>    51
<211>    4657
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    pCMVKm2.cORFopt HML-2 vector

<400>    51
gccgcggaat tcgactcta  ggccattgca tacgttgtat ctatatcata atatgtacat    60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta   120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata   180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat   240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga   300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc   360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt   420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat   480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag   540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc   600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga   660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg   720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg   780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag   840
gcacacccct ttggctctta tgcatgctat actgtttttg gcttggggcc tatacacccc   900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat   960
tattgaccac tccccctattg gtgacgatac tttccattac taatccataa catggctctt  1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga  1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac  1140
gccgtccccc gtgcccgcag ttttttattaa acatagcgtg ggatctccac gcgaatctcg  1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc  1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag  1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg  1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga  1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag  1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg gcagtgtagt ctgagcagta  1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt  1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgaacccca gcgagatgca  1680
gcgcaaggcc ccccccgcc gccgccgcca ccgcaaccgc ccccccctga cccacaagat  1740
gaacaagatg gtgaccagcg aggagcagat gaagctgccc agcaccaaga aggccgagcc  1800
ccccacctgg gcccagctga agaagctgac ccagctggcc accaagtacc tggagaacac  1860
caaggtgacc cagaccccgc gagcatgct gctggccgcc ctgatgatcg tgagcatggt  1920
gagcgccggc gtgcccaaca gcagcgagga gaccgccacc atcgagaacg gccccgctta  1980
aagaattcag actcgagcaa gtctagaaag ccatggatat cggatccact acgcgttaga  2040
gctcgctgat cagcctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc  2100
cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag  2160
gaaattgcat cgcattgtct gagtaggtgt cattctattc tggggggtgg ggtggggcag  2220
```

```
gacagcaagg gggaggattg ggaagacaat agcagggggg tgggcgaaga actccagcat    2280
gagatccccg cgctggagga tcatccagcc ggcgtcccgg aaaacgattc cgaagcccaa    2340
cctttcatag aaggcggcgg tggaatcgaa atctcgtgat ggcaggttgg gcgtcgcttg    2400
gtcggtcatt tcgaacccca gagtcccgct cagaagaact cgtcaagaag gcgatagaag    2460
gcgatgcgct gcgaatcggg agcggcgata ccgtaaagca cgaggaagcg gtcagcccat    2520
tcgccgccaa gctcttcagc aatatcacgg gtagccaacg ctatgtcctg atagcggtcc    2580
gccacaccca gccggccaca gtcgatgaat ccagaaaagc ggccattttc caccatgata    2640
ttcggcaagc aggcatcgcc atgggtcacg acgagatcct cgccgtcggg catgcgcgcc    2700
ttgagcctgg cgaacagttc ggctggcgcg agcccctgat gctcttcgtc cagatcatcc    2760
tgatcgacaa gaccggcttc catccagta cgtgctcgct cgatgcgatg tttcgcttgg    2820
tggtcgaatg ggcaggtagc cggatcaagc gtatgcagcc gccgcattgc atcagccatg    2880
atggatactt tctcggcagg agcaaggtga gatgacagga gatcctgccc cggcacttcg    2940
cccaatagca gccagtccct tcccgcttca gtgacaacgt cgagcacagc tgcgcaagga    3000
acgcccgtcg tggccagcca cgatagccgc gctgcctcgt cctgcagttc attcagggca    3060
ccggacaggt cggtcttgac aaaaagaacc gggcgcccct gcgctgacag ccggaacacg    3120
gcggcatcag agcagccgat tgtctgttgt gcccagtcat agccgaatag cctctccacc    3180
caagcggccg gagaacctgc gtgcaatcca tcttgttcaa tcatgcgaaa cgatcctcat    3240
cctgtctctt gatcagatct tgatcccctg cgccatcaga tccttggcgg caagaaagcc    3300
atccagttta ctttgcaggg cttcccaacc ttaccagagg cgcccccagc tggcaattcc    3360
ggttcgcttg ctgtccataa aaccgcccag tctagctatc gccatgtaag cccactgcaa    3420
gctacctgct ttctctttgc gcttgcgttt tcccttgtcc agatagccca gtagctgaca    3480
ttcatccggg gtcagcaccg tttctgcgga ctggctttct acgtgttccg cttcctttag    3540
cagcccttgc gccctgagtg cttgcggcag cgtgaagcta attcatggtt aaattttgt    3600
taaatcagct cattttttaa ccaataggcc gaaatcggca aaatccctta taaatcaaaa    3660
gaatagcccg agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag    3720
aacgtggact ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg ccggatcagc    3780
ttatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatc aggcgctctt    3840
ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag    3900
ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca    3960
tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt    4020
tccataggct ccgcccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc    4080
gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct    4140
ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg    4200
tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca    4260
agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta ccggtaact    4320
atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta    4380
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta    4440
actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct    4500
tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt    4560
tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga    4620
tctttttctac tgaacggtga tccccaccgg aattgcg    4657
```

```
<210>    52
<211>    4774
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    pCMVKm2.pCAP5opt HML-2 vector

<400>    52
gccgcggaat tcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat    60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta    120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata    180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat    240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga    300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc    360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt    420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat    480
gcggttttgg cagtacacca tgggcgtgg atagcggttt gactcacggg gatttccaag    540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggctttcc    600
aaaatgtcgt aataacccc ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga    660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg    720
```

Page 50

```
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg    780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag    840
gcacacccct ttggctctta tgcatgctat actgtttttg gcttggggcc tatacacccc    900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat    960
tattgaccac tcccctattg gtgacgatac tttccattac taatccataa catggctctt   1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga   1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac   1140
gccgtccccc gtgcccgcag ttttattaa acatagcgtg ggatctccac gcgaatctcg   1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc   1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag   1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg   1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga   1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag   1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg gcagtgtagt ctgagcagta   1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt   1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgaacccca gcgagatgca   1680
gcgcaaggcc cccccccgcc gccgccgcca ccgcaaccgc gcccccctga cccacaagat   1740
gaacaagatg gtgaccagcg aggagcagat gaagctgccc agcaccaaga aggccgagcc   1800
ccccacctgg gcccagctga agaagctgac ccagctggcc accaagtacc tggagaacac   1860
caaggtgacc cagacccccg agagcatgct gctggccgcc ctgatgatcg tgagcatggt   1920
ggtgtacccc accgccccca agcgccagcg ccccagccgc accggccacg acgacgcgg   1980
cggcttcgtg gagaagaagc gcggcaagtg cggcgagaag caggagcgca gcgactgcta   2040
ctgcgtgtgc gtggagcgca gccgccaccg ccgcctgcac ttcgtgctgt acgcttaaag   2100
aattcagact cgagcaagtc tagaaagcca tggatatcgg atccactacg cgttagagct   2160
cgctgatcag cctcgactgt gccttctagt tgccagccat ctgttgtttg cccctccccc   2220
gtgccttcct tgaccctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa   2280
attgcatcgc attgtctgag taggtgtcat tctattctgg ggggtggggt ggggcaggac   2340
agcaagggggg aggattggga agacaatagc aggggggtgg gcgaagaact ccagcatgag   2400
atccccgcgc tggaggatca tccagccggc gtcccggaaa acgattccga agcccaacct   2460
ttcatagaag cggcggtgg aatcgaaatc tcgtgatggc aggttgggcg tcgcttggtc   2520
ggtcatttcg aaccccagag tcccgctcag aagaactcgt caagaaggcg atagaaggcg   2580
atgcgctgcg aatcgggagc ggcgataccg taaagcacga ggaagcggtc agcccattcg   2640
ccgccaagct cttcagcaat atcacgggta gccaacgcta tgtcctgata gcggtccgcc   2700
acaccagcc ggccacagtc gatgaatcca gaaaagcggc catttttccac catgatattc   2760
ggcaagcagg catcgccatg ggtcacgacg agatcctcgc cgtcgggcat gcgcgccttg   2820
agcctggcga acagttcggc tggcgcgagc ccctgatgct cttcgtccag atcatcctga   2880
tcgacaagac cggcttccat ccgagtacgt gctcgctcga tgcgatgttt cgcttggtgg   2940
tcgaatgggc aggtagccgg atcaagcgta tgcagccgcc gcattgcatc agccatgatg   3000
gatactttct cggcaggagc aaggtgagat gacaggagat cctgccccgg cacttcgccc   3060
aatagcagcc agtcccttcc cgcttcagtg acaacgtcga gcacagctgc gcaaggaacg   3120
cccgtcgtgg ccagccacga tagccgcgct gcctcgtcct gcagttcatt cagggcaccg   3180
gacaggtcgg tcttgacaaa aagaaccggg cgcccctgcg ctgacagccg gaacacggcg   3240
gcatcagagc agccgattgt ctgttgtgcc cagtcatagc cgaatagcct ctccacccaa   3300
gcggccggag aacctgcgtg caatccatct tgttcaatca tgcgaaacga tcctcatcct   3360
gtctcttgat cagatcttga tcccctgcgc catcagatcc ttggcggcaa gaaagccatc   3420
cagtttactt tgcagggctt cccaacctta ccagagggcg ccccagctgg caattccggt   3480
tcgcttgctg tccataaaac cgcccagtct agctatcgcc atgtaagccc actgcaagct   3540
acctgctttc tctttgcgct tgcgttttcc cttgtccaga tagcccagta gctgacattc   3600
atccggggtc agcaccgttt ctgcggactg gctttctacg tgttccgctt cctttagcag   3660
cccttgcgcc ctgagtgctt gcggcagcgt gaagctaatt catggttaaa tttttgttaa   3720
atcagctcat tttttaacca ataggccgaa atcggcaaaa tcccttataa atcaaaagaa   3780
tagcccgaga tagggttgag tgttgttcca gtttggaaca agagtccact attaaagaac   3840
gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccg gatcagctta   3900
tgcggtgtga ataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcttccg   3960
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc   4020
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga agaacatgt   4080
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc   4140
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa   4200
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc   4260
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg   4320
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc   4380
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc   4440
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca   4500
```

Page 51

```
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact   4560
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg   4620
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt   4680
ttgtttgcaa gcagcagatt acgcgcagaa aaaaggatc tcaagaagat cctttgatct    4740
tttctactga acggtgatcc ccaccggaat tgcg                               4774
```

<210> 53
<211> 6483
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.gag wt PCAV vector

<400> 53
```
gccgcggaat tcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat    60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta   120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata   180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat   240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga   300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc   360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt   420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat   480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag   540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc   600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga   660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg   720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg   780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag   840
gcacacccct ttggctctta tgcatgctat actgtttttg cttggggcc tatacacccc     900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat   960
tattgaccac tcccctattg gtgacgatac tttccattac taatccataa catggctctt   1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga   1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac   1140
gccgtccccc gtgcccgcag tttttattaa acatagcgtg ggatctccac gcgaatctcg   1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc   1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag   1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg   1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga   1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag   1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg cagtgtagt ctgagcagta     1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt   1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atggggcaaa ctgaaagtaa   1680
atatgcctct tatctcagct ttattaaaat tctttaaga agaggggag ttagagcttc      1740
tacagaaaat ctaattacgc tatttcaaac aatagaacaa ttctgcccat ggtttccaga   1800
acagggaact ttagatctaa aagattggga aaaaattggc aaagaattaa aacaagcaaa   1860
tagggaaggt aaaatcatcc cacttacagt atggaatgat tgggccatta ttaaagcaac   1920
tttagaacca tttcaaacag gagaagatat tgtttcagtt tctgatgccc ctaaaagctg   1980
tgtaacagat tgtgaagaag aggcagggac agaatcccag caaggaacgg aaagttcaca   2040
ttgtaaatat gtagcagagt ctgtaatggc tcagtcaacg caaaatgttg actacagtca   2100
attacaggag ataatatacc ctgaatcatc aaaattgggg gaaggaggtc agaatcatt    2160
ggggccatca gagcctaaac cacgatcgcc atcaactcct cctcccgtgg ttcagatgcc   2220
tgtaacatta caacctcaaa cgcaggttag acaagcacaa accccaagag aaaatcaagt   2280
agaaagggac agagtctcta tcccggcaat gccaactcag atacagtatc cacaatatca   2340
gccggtagaa aataagaccc aaccgctggt agtttatcaa taccggctgc caaccgagct   2400
tcagtatcgg cctccttcag aggttcaata cagacctcaa gcggtgtgtc ctgtgccaaa   2460
tagcacggca ccataccagc aacccacagc gatggcgtct aattcaccag caacacagga   2520
cgcggcgctg tatcctcagc cgcccactgt gagacttaat cctacagcat cacgtagtgg   2580
acagggtggt gcactgcatg cagtcattga tgaagccaga aaacagggcg atcttgaggc   2640
atggcggttc ctggtaattt tacaactggt acaggccggg aagagactc aagtaggagc      2700
gcctgcccga gctgagacta gatgtgaacc tttcaccatg aaaatgttaa aagatataaa   2760
ggaaggagtt aaacaatatg gatccaactc cccttatata agaacattat tagattccat   2820
tgctcatgga aatagactta ctccttatga ctgggaaatt ttggccaaat cttccctttc   2880
```
                                  Page 52

```
atcctctcag tatctacagt ttaaaacctg gtggattgat ggagtacaag aacaggtacg 2940
aaaaaatcag gctactaagc ccactgttaa tatagacgca gaccaattgt taggaacagg 3000
tccaaattgg agcaccatta accaacaatc agtgatgcag aatgaggcta ttgaacaagt 3060
aagggctatt tgcctcaggg cctggggaaa aattcaggac ccaggaacag ctttccctat 3120
taattcaatt agacaaggct ctaaagagcc atatcctgac tttgtggcaa gattacaaga 3180
tgctgctcaa aagtctatta cagatgacaa tgcccgaaaa gttattgtag aattaatggc 3240
ctatgaaaat gcaaatccag aatgtcagtc ggccataaag ccattaaaag gaaaagttcc 3300
agcaggagtt gatgtaatta cagaatatgt gaaggcttgt gatgggattg gaggagctat 3360
gcataaggca atgctaatgg ctcaagcaat gaggggggctc actctaggag gacaagttag 3420
aacatttggg aaaaaatgtt ataattgtgg tcaaatcggt catctgaaaa ggagttgccc 3480
agtcttaaat aaacagaata taataaatca agctattaca gcaaaaaata aaaagccatc 3540
tggcctgtgt ccaaaatgtg gaaaaggaaa acattgggcc aatcaatgtc attctaaatt 3600
tgataaggat gggcaaccat tgtcgggaaa caggaagagg ggccagcctc aggcccccca 3660
acaaactggg gcattcccag ttcaactgtt tgttcctcag ggttttcaag gacaacaacc 3720
cctacagaaa ataccaccac ttcagggagt cagccaatta caacaatcca acagctgtcc 3780
cgcgccacag caggcagcac cgcagtaaga attcagactc gagcaagtct agaaagccat 3840
ggatatcgga tccactacgc gttagagctc gctgatcagc ctcgactgtg ccttctagtt 3900
gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 3960
ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 4020
ctattctggg gggtggggtg gggcaggaca gcaagggggga ggattgggaa gacaatagca 4080
ggggggtggg cgaagaactc cagcatgaga tccccgcgct ggaggatcat ccagccggcg 4140
tcccggaaaa cgattccgaa gcccaacctt tcatagaagg cggcggtgga atcgaaatct 4200
cgtgatggca ggttgggcgt cgcttggtcg gtcatttcga accccagagt cccgctcaga 4260
agaactcgtc aagaaggcga tagaaggcga tgcgctgcga atcgggagcg gcgataccgt 4320
aaagcacgag gaagcggtca gcccattcgc cgccaagctc ttcagcaata tcacgggtag 4380
ccaacgctat gtcctgatag cggtccgcca cacccagccg gccacagtcg atgaatccag 4440
aaaagcggcc attttccacc atgatattcg gcaagcaggc atcgccatgg gtcacgacga 4500
gatcctcgcc gtcgggcatg cgcgccttga gcctggcgaa cagttcggct ggcgcgagcc 4560
cctgatgctc ttcgtccaga tcatcctgat cgacaagacc ggcttccatc cgagtacgtg 4620
ctcgctcgat gcgatgtttc gcttggtggt cgaatgggca ggtagccgga tcaagcgtat 4680
gcagccgccg cattgcatca gccatgatgg atactttctc ggcaggagca aggtgagatg 4740
acaggagatc ctgccccggc acttcgccca atagcagcca gtcccttccc gcttcagtga 4800
caacgtcgag cacagctgcg caaggaacgc ccgtcgtggc cagccacgat agccgcgctg 4860
cctcgtcctg cagttcattc agggcaccgg acaggtcggt cttgacaaaa agaaccgggc 4920
gcccctgcgc tgacagccgg aacacggcgg catcagagca gccgattgtc tgttgtgccc 4980
agtcatagcc gaatagcctc tccacccaag cggccggaga acctgcgtgc aatccatctt 5040
gttcaatcat gcgaaacgat cctcatcctg tctcttgatc agatcttgat cccctgcgcc 5100
atcagatcct tggcggcaag aaagccatcc agtttacttt gcagggcttc ccaaccttac 5160
cagagggcgc cccagctggc aattccggtt cgcttgctgt ccataaaacc gcccagtcta 5220
gctatcgcca tgtaagccca ctgcaagcta cctgctttct ctttgcgctt gcgttttccc 5280
ttgtccagat agcccagtag ctgacattca tccggggtca gcaccgtttc tgcggactgg 5340
ctttctacgt gttccgcttc ctttagcagc ccttgcgccc tgagtgcttg cggcagcgtg 5400
aagctaattc atggttaaat ttttgttaaa tcagctcatt ttttaaccaa taggccgaaa 5460
tcggcaaaat cccttataaa tcaaaagaat agcccgagat agggttgagt gttgttccag 5520
tttggaacaa gagtccacta ttaaagaacg tggactccaa cgtcaaaggg cgaaaaaccg 5580
tctatcaggg cgatggccgg atcagcttat gcggtgtgaa ataccgcaca gatgcgtaag 5640
gagaaaatac cgcatcaggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt 5700
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt tatccacaga 5760
atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 5820
taaaaaggcc gcgttgctgg cgttttttcca taggctccgc cccctgacg agcatcacaa 5880
aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt 5940
tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct 6000
gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct 6060
cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 6120
cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt 6180
atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc 6240
tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat 6300
ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa 6360
acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa 6420
aaaaggatct caagaagatc ctttgatctt ttctactgaa cggtgatccc caccggaatt 6480
gcg.                                                                6483
```
<210> 54
```

<211> 6340
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.gagopt HML-2 vector

<400> 54
```
gccgcggaat ttcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat   60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta  120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata  180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat  240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga  300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc  360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt  420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat  480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag  540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggctttcc  600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga  660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg  720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccggcggc gggaacggtg  780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag  840
gcacacccct ttggctctta tgcatgctat actgtttttg gcttggggcc tatacacccc  900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat  960
tattgaccac tcccctattg gtgacgatac tttccattac taatccataa catggctctt 1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga 1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac 1140
gccgtccccc gtgcccgcag ttttattaa acatagcgtg ggatctccac gcgaatctcg 1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc 1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag 1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg 1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga 1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag 1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg gcagtgtagt ctgagcagta 1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt 1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgggccaga ccaagagcaa 1680
gatcaagagc aagtacgcca gctacctgag cttcatcaag atcctgctga gcgcggcgg 1740
cgtgaaggtg agcaccaaga acctgatcaa gctgttccag atcatcgagc agttctgccc 1800
ctggttcccc gagcagggca ccctggacct gaaggactgg aagcgcatcg gcaaggagct 1860
gaagcaggcc ggccgcaagg gcaacatcat ccccctgacc gtgtggaacg actgggccat 1920
catcaaggcc gccctggaca ccttccagac cgaggaggac agcgtgagcg tgagcgacgc 1980
ccccggcagc tgcatcatcg actgcaacga gaacacccgc aagaagagcc agaaggagac 2040
cgagggcctg cactgcgagt acgtggccga gcccgtgatg gcccagaaca cccagaacgt 2100
ggactacaac cagctgcagg aggtgatcta ccccgagacc ctgaagctgg agggcaaggg 2160
ccccgagctg gtgggcccca gcgagagcaa gccccgcggc accagccccc tgcccgccgg 2220
ccaggtgccc gtgaccctgc agccccagaa gcaggtgaag gagaacaaga cccagccccc 2280
cgtggcctac cagtactggc ccccgccga gctgcagtac cgcccccccc cgagagcca 2340
gtacggctac cccggcatgc cccccgcccc ccaggccgc gcccctacc cccagccccc 2400
cacccgccgc ctgaacccca ccgcccccc cagccgccag ggcagcaagc tgcacgagat 2460
catcgacaag agccgcaagg agggcgacac cgaggcctgg cagttccccg tgaccctgga 2520
gcccatgccc cccggcgagg gcgcccagga gggcgagccc cccaccgtgg aggcccgcta 2580
caagagcttc agcatcaaga agctgaagga catgaaggag ggcgtgaagc agtacggccc 2640
caacagcccc tacatgcgca ccctgctgga cagcatcgcc acggccacc gcctgatccc 2700
ctacgactgg gagatcctgg ccaagagcag cctgagcccc agccagttcc tgcagttcaa 2760
gacctggtgg atcgacggcg tgcaggagca ggtcgccgc aaccgcgccg ccaacccccc 2820
cgtgaacatc gacgccgacc agctgctggg catcggccag aactggagca ccatcagcca 2880
gcaggccctg atgcagaacg aggccatcga gcaggtgcgc gccatctgcc tgcgcgcctg 2940
ggagaagatc caggaccccg gcagcacctg ccccagcttc aacaccgtgc gccagggcag 3000
caaggagccc taccccgact tcgtggcccg cctgcaggac gtggcccaga gagcatcgc 3060
cgacgagaag gcccgcaagg tgatcgtgga gctgatggcc tacgagaacg ccaacccga 3120
gtgccagagc gccatcaagc ccctgaaggg caaggtgccc gccggcagcg acgtgatcag 3180
cgagtacgtg aaggcctgcg acggcatcgg cggcgccatg cacaaggcca tgctgatggc 3240
ccaggccatc accggcgtgg tgctgggcgg ccaggtgcgc accttcggcc gcaagtgcta 3300
```

```
caactgcggc cagatcggcc acctgaagaa gaactgcccc gtgctgaaca agcagaacat    3360
caccatccag gccaccacca ccggccgcga gcccccgac ctgtgcccc gctgcaagaa     3420
gggcaagcac tgggccagcc agtgccgcag caagttcgac aagaacggcc agcccctgag   3480
cggcaacgag cagcgcggcc agcccaggc ccccagcag accggcgcct tccccatcca     3540
gcccttcgtg ccccagggct tccagggcca gcagccccc ctgagccagg tgttccaggg    3600
catcagccag ctgccccagt acaacaactg ccccccccc caggccgccg tgcagcaggc    3660
ttaaagaatt cagactcgag caagtctaga aagccatgga tatcggatcc actacgcgtt   3720
agagctcgct gatcagcctc gactgtgcct tctagttgcc agccatctgt tgtttgcccc   3780
tcccccgtgc cttccttgac cctggaaggt gccactccca ctgtcctttc ctaataaaat   3840
gaggaaattg catcgcattg tctgagtagg tgtcattcta ttctggggg tgggtgggg     3900
caggacagca aggggagga ttgggaagac aatagcaggg gggtgggcga agaactccag     3960
catgagatcc ccgcgctgga ggatcatcca gccggcgtcc cggaaaacga ttccgaagcc   4020
caacctttca tagaaggcgg cggtggaatc gaaatctcgt gatggcaggt tgggcgtcgc   4080
ttggtcggtc atttcgaacc ccagagtccc gctcagaaga actcgtcaag aaggcgatag   4140
aaggcgatgc gctgcgaatc gggagcggcg ataccgtaaa gcacgaggaa gcggtcagcc   4200
cattcgccgc caagctcttc agcaatatca cgggtagcca acgctatgtc ctgatagcgg   4260
tccgccacac ccagccggcc acagtcgatg aatccagaaa agcggccatt ttccaccatg   4320
atattcggca agcaggcatc gccatgggtc acgacgagat cctcgccgtc gggcatgcgc   4380
gccttgagcc tggcgaacag ttcggctggc gcgagcccct gatgctcttc gtccagatca   4440
tcctgatcga caagaccggc ttccatccga gtacgtgctc gctcgatgcg atgtttcgct   4500
tggtggtcga atgggcaggt agccggatca agcgtatgca gccgccgcat tgcatcagcc   4560
atgatggata ctttctcggc aggagcaagg tgagatgaca ggagatcctg ccccggcact   4620
tcgcccaata gcagccagtc ccttcccgct tcagtgacaa cgtcgagcac agctgcgcaa   4680
ggaacgcccg tcgtggccag ccacgatagc cgcgctgcct cgtcctgcag ttcattcagg   4740
gcaccggaca ggtcggtctt gacaaaaaga accgggcgcc cctgcgctga cagccggaac   4800
acggcggcat cagagcagcc gattgtctgt tgtgcccagt catagccgaa tagcctctcc   4860
acccaagcgg ccggagaacc tgcgtgcaat ccatcttgtt caatcatgcg aaacgatcct   4920
catcctgtct cttgatcaga tcttgatccc ctgcgccatc agatccttgg cggcaagaaa   4980
gccatccagt ttactttgca gggcttccca accttaccag agggcgcccc agctggcaat   5040
tccggttcgc ttgctgtcca taaaaccgcc cagtctagct atcgccatgt aagcccactg   5100
caagctacct gctttctctt tgcgcttgcg ttttcccttg tccagatagc ccagtagctg   5160
acattcatcc ggggtcagca ccgtttctgc ggactggctt tctacgtgtt ccgcttcctt   5220
tagcagccct tgcgccctga gtgcttgcgg cagcgtgaag ctaattcatg gttaaatttt   5280
tgttaaatca gctcattttt taaccaatag gccgaaatcg gcaaaatccc ttataaatca   5340
aaagaatagc ccgagatagg gttgagtgtt gttccagttt ggaacaagag tccactatta   5400
aagaacgtgg actccaacgt caaagggcga aaaccgtct atcagggcga tggccggatc    5460
agcttatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgct   5520
cttccgcttc ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat   5580
cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga   5640
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt   5700
ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt   5760
ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc   5820
gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa   5880
gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct   5940
ccaagctggg ctgtgtgcac gaacccccg ttcagcccga ccgctgcgcc ttatccggta    6000
actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg   6060
gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc   6120
ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta   6180
ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg   6240
gtttttttgt ttgcaagcag cagattacgc gcagaaaaa aggatctcaa gaagatcctt    6300
tgatctttc tactgaacgg tgatccccac cggaattgcg                          6340
```

<210>    55
<211>    5344
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    pCMVKm2.Protopt HML-2 vector

<400>    55
```
gccgcggaat tcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat     60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta    120
```

```
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata    180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat    240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga    300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc    360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt    420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat    480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag    540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc    600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga    660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg    720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg    780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag    840
gcacacccct ttggctctta tgcatgctat actgtttttg gcttggggcc tatacacccc    900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat    960
tattgaccac tcccctattg gtgacgatac tttccattac taatccataa catggctctt   1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga   1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac   1140
gccgtccccc gtgcccgcag tttttattaa acatagcgtg ggatctccac gcgaatctcg   1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc   1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag   1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg   1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga   1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag   1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg gcagtgtagt ctgagcagta   1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt   1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgtgggcca ccatcgtggg   1680
caagcgcgcc aagggccccg ccagcggccc caccaccaac tggggcatcc ccaacagcgc   1740
catctgcagc agcggcttca gcggcaccac cacccccacc gtgcccagcg tgagcggcaa   1800
caagcccgtg accaccatcc agcagctgag ccccgccacc agcggcagcg ccgccgtgga   1860
cctgtgcacc atccaggccg tgagcctgct gcccggcgag cccccccaga agacccccac   1920
cggcgtgtac ggccccctgc ccaagggcac cgtgggcctg atcctgggcc gcagcagcct   1980
gaacctgaag ggcgtgcaga tccacaccag cgtggtggac agcgactaca agggcgagat   2040
ccagctggtg atcagcagca gcatcccctg gagcgccagc ccccgcgacc gcatcgccca   2100
gctgctgctg ctgccctaca tcaagggcgg caacagcgag atcaagcgca tcggcggcct   2160
gggcagcacc gaccccaccg gcaaggccgc ctactgggcc agccaggtga gcgagaaccg   2220
ccccgtgtgc aaggccatca tccagggcaa gcagttcgag ggcctggtgg acaccggcgc   2280
cgacgtgagc atcatcgccc tgaaccagtg gcccaagaac tggcccaagc agaaggccgt   2340
gaccggcctg gtgggcatcg gcaccgccag cgaggtgtac cagagcaccg agatcctgca   2400
ctgcctgggc cccgacaacc aggagagcac cgtgcagccc atgatcacca gcatccccct   2460
gaacctgtgg ggccgcgacc tgctgcagca gtggggcgcc gagatcacca tgcccgcccc   2520
cagctacagc cccaccagcc agaagatcat gaccaagatg ggctacatcc ccggcaaggg   2580
cctgggcaag aacgaggacg gcatcaagat ccccgtggag gccaagatca accaggagcg   2640
cgagggcatc ggcaacccct gcgcttaaag aattcagact cgagcaagtc tagaaagcca   2700
tggatatcgg atccactacg cgttagagct cgctgatcag cctcgactgt gccttctagt   2760
tgccagccat ctgttgtttg cccctccccc gtgccttcct tgaccctgga aggtgccact   2820
cccactgtcc tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat   2880
tctattctgg ggggtggggt ggggcaggac agcaaggggg aggattggga agacaatagc   2940
aggggggtgg gcgaagaact ccagcatgag atccccgcgc tggaggatca tccagccggc   3000
gtcccggaaa acgattccga agcccaacct ttcatagaag cggcggtgg aatcgaaatc   3060
tcgtgatggc aggttgggcg tcgcttggtc ggtcatttcg aaccccagag tcccgctcag   3120
aagaactcgt caagaaggcg atagaaggcg atgcgctgcg aatcgggagc ggcgataccg   3180
taaagcacga ggaagcggtc agcccattcg ccgccaagct cttcagcaat atcacgggta   3240
gccaacgcta tgtcctgata gcggtccgcc acacccagcc ggccacagtc gatgaatcca   3300
gaaaagcggc cattttccac catgatattc ggcaagcagg catcgccatg ggtcacgacg   3360
agatcctcgc cgtcgggcat gcgcgccttg agcctggcga acagttcggc tggcgcgagc   3420
ccctgatgct cttcgtccag atcatcctga tcgacaagac cggcttccat ccgagtacgt   3480
gctcgctcga tgcgatgttt cgcttggtgg tcgaatgggc aggtagccgg atcaagcgta   3540
tgcagccgcc gcattgcatc agccatgatg gatactttct cggcaggagc aaggtgagat   3600
gacaggagat cctgccccgg cacttcgccc aatagcagcc agtcccttcc cgcttcagtg   3660
acaacgtcga gcacagctgc gcaaggaacg cccgtcgtgg ccagcacga tagccgcgct   3720
gcctcgtcct gcagttcatt cagggcaccg gacaggtcgg tcttgacaaa aagaaccggg   3780
cgcccctgcg ctgacagccg gaacacggcg gcatcagagc agccgattgt ctgttgtgcc   3840
cagtcatagc cgaatagcct ctccacccaa gcggccggag aacctgcgtg caatccatct   3900
```

```
tgttcaatca tgcgaaacga tcctcatcct gtctcttgat cagatcttga tcccctgcgc   3960
catcagatcc ttggcggcaa gaaagccatc cagtttactt tgcagggctt cccaacctta   4020
ccagagggcg ccccagctgg caattccggt tcgcttgctg tccataaaac cgcccagtct   4080
agctatcgcc atgtaagccc actgcaagct acctgctttc tctttgcgct tgcgttttcc   4140
cttgtccaga tagcccagta gctgacattc atccggggtc agcaccgttt ctgcggactg   4200
gctttctacg tgttccgctt cctttagcag cccttgcgcc ctgagtgctt gcggcagcgt   4260
gaagctaatt catggttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa   4320
atcggcaaaa tcccttataa atcaaaagaa tagcccgaga tagggttgag tgttgttcca   4380
gtttggaaca agagtccact attaaagaac gtggactcca acgtcaaagg cgaaaaacc    4440
gtctatcagg gcgatggccg gatcagctta tgcggtgtga ataccgcac agatgcgtaa     4500
ggagaaaata ccgcatcagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg   4560
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag   4620
aatcagggga taacgcagga agaacatgt gagcaaaagg ccagcaaaag gccaggaacc    4680
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccctgac gagcatcaca     4740
aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt   4800
ttcccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc    4860
tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc   4920
tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc   4980
ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact   5040
tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg   5100
ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggta   5160
tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca   5220
aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa   5280
aaaaaggatc tcaagaagat cctttgatct tttctactga acggtgatcc ccaccggaat   5340
tgcg                                                                5344
```

<210>   56
<211>   7211
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   pCMVKm2.Polopt HML-2 vector

<400>   56
```
gccgcggaat tcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat    60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta   120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata   180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat   240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga   300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc   360
ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt acatgacctt   420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat   480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag   540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc   600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga   660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg   720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg   780
cattggaacg cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag   840
gcacacccct ttggctctta tgcatgctat actgtttttg gcttggggcc tatacacccc   900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat   960
tattgaccac tccctattg gtgacgatac tttccattac taatccataa catggctctt   1020
tgccacaact atctctattg ctatatgcc aatactctgt ccttcagaga ctgacacgga   1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac  1140
gccgtccccc gtgccgcag ttttattaa acatagcgtg ggatctccac gcgaatctcg    1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc   1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag   1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg   1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga   1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag   1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg cagtgtagt ctgagcagta    1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt   1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgaacaaga gccgcaagcg   1680
```
                              Page 57

```
ccgcaaccgc gagagcctgc tgggcgccgc caccgtggag cccccaagc ccatcccct   1740
gacctggaag accgagaagc ccgtgtgggt gaaccagtgg cccctgccca agcagaagct   1800
ggaggccctg cacctgctgg ccaacgagca gctggagaag ggccacatcg agcccagctt   1860
cagcccctgg aacagccccg tgttcgtgat ccagaagaag agcggcaagt ggcgcatgct   1920
gaccgacctg cgcgccgtga acgccgtgat ccagcccatg ggccccctgc agcccggcct   1980
gcccagcccc gccatgatcc ccaaggactg gcccctgatc atcatcgacc tgaaggactg   2040
cttcttcacc atcccccctgg ccgagcagga ctgcgagaag ttcgccttca ccatccccgc   2100
catcaacaac aaggagcccg ccacccgctt ccagtggaag gtgctgcccc agggcatgct   2160
gaacagcccc accatctgcc agaccttcgt gggccgcgcc ctgcagcccg tgcgcgagaa   2220
gttcagcgac tgctacatca tccactgcat cgacgacatc ctgtgcgccg ccgagaccaa   2280
ggacaagctg atcgactgct acaccttcct gcaggccgag gtggccaacg ccggcctggc   2340
catcgccagc gacaagatcc agaccagcac ccccttccac tacctgggca tgcagatcga   2400
gaaccgcaag atcaagcccc agaagatcga gatccgcaag gacaccctga agaccctgaa   2460
cgacttccag aagctgctgg gcgacatcaa ctggatccgc cccaccctgg gcatccccac   2520
ctacgccatg agcaacctgt tcagcatcct gcgcggcgac agcgacctga acagcaagcg   2580
catgctgacc cccgaggcca ccaaggagat caagctggtg gaggagaaga tccagagcgc   2640
ccagatcaac cgcatcgacc ccctggcccc cctgcagctg ctgatcttcg ccaccgccca   2700
cagccccacc ggcatcatca tccagaacac cgacctggtg gagtggagct tcctgcccca   2760
cagcaccgtg aagaccttca ccctgtacct ggaccagatc gccaccctga tcggccagac   2820
ccgcctgcgc atcatcaagc tgtgcggcaa cgaccccgac aagatcgtgg tgcccctgac   2880
caaggagcag gtgcgccagg ccttcatcaa cagcggccgc tggaagatcg gcctggccaa   2940
cttcgtgggc atcatcgaca accactaccc caagaccaag atcttccagt tcctgaagct   3000
gaccacctgg atcctgccca agatcacccg ccgcgagccc ctggagaacg ccctgaccgt   3060
gttcaccgac ggcagcagca cggcaaggc cgcctacacc ggccccaagg agcgcgtgat   3120
caagacccc taccagagcg cccagcgcgc cgagctggtg gccgtgatca ccgtgctgca   3180
ggacttcgac cagcccatca acatcatcag cgacagcgcc tacgtggtgc aggccacccg   3240
cgacgtggag accgccctga tcaagtacag catggacgac cagctgaacc agctgttcaa   3300
cctgctgcag cagaccgtgc gcaagcgcaa cttcccctc tacatcaccc acatccgcgc   3360
ccacaccaac ctgcccggcc ccctgaccaa ggccaacgag caggccgacc tgctggtgag   3420
cagcgccctg atcaaggccc aggagctgca cgccctgacc cacgtgaacg ccgccggcct   3480
gaagaacaag ttcgacgtga cctggaagca ggccaaggac atcgtgcagc actgcaccca   3540
gtgccaggtg ctgcacctgc ccacccagga ggccggcgtg aaccccccgcg cctgtgccc   3600
caacgccctg tggcagatgg acgtgaccca cgtgcccagc ttcggccgcc tgagctacgt   3660
gcacgtgacc gtggacacct acagccactt catctgggcc acctgccaga ccggcgagag   3720
caccagccac gtgaagaagc acctgctgag ctgcttcgcc gtgatgggcg tgcccgagaa   3780
gatcaagacc gacaacggcc ccggctactg cagcaaggcc ttccagaagt tcctgagcca   3840
gtggaagatc agccacacca ccggcatccc ctacaacagc cagggccagg ccatcgtgga   3900
gcgcaccaac cgcaccctga gacccagct ggtgaagcag aaggagggcg cgacagcaa   3960
ggagtgcacc accccccaga tgcagctgaa cctggccctg tacaccctga acttcctgaa   4020
catctaccgc aaccagacca ccaccagcgc cgagcagcac ctgaccggca agaagaacag   4080
cccccacgag ggcaagctga tctggtggaa ggacaacaag aacaagacct gggagatcgg   4140
caaggtgatc acctggggcc gcggcttcgc ctgcgtgagc cccggcgaga accagctgcc   4200
cgtgtggatc cccacccgcc acctgaagtt ctacaacgag cccatccgcg acgccaagaa   4260
gagcaccagc gccgagaccg agaccgacca gagcagcacc gtggacagcc aggacgagca   4320
gaacggcgac gtgcgccgca ccgacgaggt ggccatccac caggagggcc gcgccgccaa   4380
cctgggcacc accaaggagg ccgacgccgt gagctacaag atcagccgcg agcacaaggg   4440
cgacaccaac ccccgcgagt acgccgcctg cagcctggac gactgcatca cggcggcaa   4500
gagcccctac gcctgccgca gcagctgcag cttaaagaat tcagactcga gcaagtctag   4560
aaagccatgg atatcggatc cactacgcgt tagagctcgc tgatcagcct cgactgtgcc   4620
ttctagttgc cagccatctg ttgtttgccc ctccccgtg ccttccttga ccctggaagg   4680
tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag   4740
gtgtcattct attctggggg gtggggtggg gcaggacagc aaggggggagg attgggaaga   4800
caatagcagg ggggtgggcg aagaactcca gcatgagatc cccgcgctgg aggatcatcc   4860
agccggcgtc ccggaaaacg attccgaagc ccaacctttc atagaaggcg gcggtggaat   4920
cgaaatctcg tgatggcagg ttgggcgtcg cttggtcggt catttcgaac cccagagtcc   4980
cgctcagaag aactcgtcaa gaaggcgata gaaggcgatg cgctgcgaat cgggagcggc   5040
gataccgtaa agcacgagga agcggtcagc ccattcgccg ccaagctctt cagcaatatc   5100
acgggtagcc aacgctatgt cctgatagcg gtccgccaca cccagccggc cacagtcgat   5160
gaatccagaa aagcggccat tttccaccat gatattcggc aagcaggcat cgccatgggt   5220
cacgacgaga tcctcgccgt cgggcatgcg cgccttgagc ctggcgaaca gttcggctgg   5280
cgcgagcccc tgatgctctt cgtccagatc atcctgatcg acaagaccgg cttccatccg   5340
agtacgtgct cgctcgatgc gatgtttcgc ttggtggtcg aatgggcagg tagccggatc   5400
aagcgtatgc agccgccgca ttgcatcagc catgatggat actttctcgg caggagcaag   5460
```

```
gtgagatgac aggagatcct gccccggcac ttcgcccaat agcagccagt cccttcccgc   5520
ttcagtgaca acgtcgagca cagctgcgca aggaacgccc gtcgtggcca gccacgatag   5580
ccgcgctgcc tcgtcctgca gttcattcag ggcaccggac aggtcggtct tgacaaaaag   5640
aaccgggcgc ccctgcgctg acagccggaa cacggcggca tcagagcagc cgattgtctg   5700
ttgtgcccag tcatagccga atagcctctc cacccaagcg gccggagaac ctgcgtgcaa   5760
tccatcttgt tcaatcatgc gaaacgatcc tcatcctgtc tcttgatcag atcttgatcc   5820
cctgcgccat cagatccttg gcggcaagaa agccatccag tttactttgc agggcttccc   5880
aaccttacca gagggcgccc cagctggcaa ttccggttcg cttgctgtcc ataaaaccgc   5940
ccagtctagc tatcgccatg taagcccact gcaagctacc tgctttctct ttgcgcttgc   6000
gttttccctt gtccagatag cccagtagct gacattcatc cggggtcagc accgtttctg   6060
cggactggct ttctacgtgt tccgcttcct ttagcagccc ttgcgccctg agtgcttgcg   6120
gcagcgtgaa gctaattcat ggttaaattt ttgttaaatc agctcatttt ttaaccaata   6180
ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag cccgagatag ggttgagtgt   6240
tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg   6300
aaaaaccgtc tatcagggcg atggccggat cagcttatgc ggtgtgaaat accgcacaga   6360
tgcgtaagga gaaaataccg catcaggcgc tcttccgctt cctcgctcac tgactcgctg   6420
cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta   6480.
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc   6540
aggaaccgta aaaaggccgc gttgctggcg ttttttccata ggctccgccc ccctgacgag   6600
catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac   6660
caggcgtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc   6720
ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt   6780
aggtatctca gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccccc   6840
gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga   6900
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta   6960
ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta   7020
tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga   7080
tccggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg   7140
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctactgaacg gtgatcccca   7200
ccggaattgc g                                                        7211
```

```
<210>    57
<211>    318
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    57
atgaacccat cagagatgca aagaaaagca cctccgcgga gacggagaca tcgcaatcga    60
gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag aagaacagat gaagttgcca   120
tccaccaaga aggcagagcc gccaacttgg gcacaactaa agaagctgac gcagttagct   180
acaaaatatc tagagaacac aaaggtgaca caaaccccag agagtatgct gcttgcagcc   240
ttgatgattg tatcaatggt gtctgcaggt gtacccaaca gctccgaaga gacagcgacc   300
atcgagaacg ggccatga                                                 318

<210>    58
<211>    321
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Modified cORF sequence

<400>    58
atgaacccca gcgagatgca gcgcaaggcc ccccccgcc gccgccgcca ccgcaaccgc    60
gcccccctga cccacaagat gaacaagatg gtgaccagcg aggagcagat gaagctgccc   120
agcaccaaga aggccgagcc ccccacctgg gcccagctga agaagctgac ccagctggcc   180
accaagtacc tggagaacac caaggtgacc cagacccccg agagcatgct gctggccgcc   240
ctgatgatcg tgagcatggt gagcgccggc gtgcccaaca gcagcgagga gaccgccacc   300
atcgagaacg gccccgctta a                                             321

<210>    59
<211>    435
<212>    DNA
```

<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    59
atgaacccat cggagatgca aagaaaagca cctccgcgga gacggagaca tcgcaatcga    60
gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag aagaacagat gaagttgcca   120
tccaccaaga aggcagagcc gccaacttgg gcacaactaa agaagctgac gcagttagct   180
acaaaatatc tagagaacac aaaggtgaca caaaccccag agagtatgct gcttgcagcc   240
ttgatgattg tatcaatggt ggtgtaccca acagctccga agagacagcg accatcgaga   300
acgggccatg atgacgatgg cggttttgtc gaaaagaaaa gggggaaatg tggggaaaag   360
caagagagat cagattgtta ctgtgtctgt gtagaagaa gtagacatag gagactccat   420
tttgttctgt actaa                                                     435

<210>    60
<211>    438
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Modified PCAP5 sequence

<400>    60
atgaaccca gcgagatgca gcgcaaggcc ccccccgcc gccgccgcca ccgcaaccgc    60
gccccctga cccacaagat gaacaagatg gtgaccagcg aggagcagat gaagctgccc   120
agcaccaaga aggccgagcc ccccacctgg gcccagctga agaagctgac ccagctggcc   180
accaagtacc tggagaacac caaggtgacc cagacccccg agagcatgct gctggccgcc   240
ctgatgatcg tgagcatggt ggtgtacccc accgcccccca agcgccagcg ccccagccgc   300
accggccacg acgacgacgg cggcttcgtg gagaagaagc gcggcaagtg cggcgagaag   360
caggagcgca gcgactgcta ctgcgtgtgc gtggagcgca gccgccaccg ccgcctgcac   420
ttcgtgctgt acgcttaa                                                  .438

<210>    61
<211>    2001
<212>    DNA
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    61
atggggcaaa ctaaaagtaa aattaaaagt aaatatgcct cttatctcag ctttattaaa    60
attcttttaa aaagagggg agttaaagta tctacaaaaa atctaatcaa gctatttcaa   120
ataatagaac aatttttgccc atggtttcca gaacaaggaa ctttagatct aaaagattgg   180
aaaagaattg gtaaggaact aaaacaagca ggtaggaagg gtaatatcat tccacttaca   240
gtatggaatg attgggccat tattaaagca gctttagaac catttcaaac agaagaagat   300
agcgtttcag tttctgatgc ccctggaagc tgtataatag attgtaatga aaacacaagg   360
aaaaaatccc agaaagaaac ggaaggttta cattgcgaat atgtagcaaga gccggtaatg   420
gctcagtcaa cgcaaaatgt tgactataat caattacagg aggtgatata tcctgaaacg   480
ttaaaattag aaggaaaagg tccagaatta gtggggccat cagagtctaa accacgaggc   540
acaagtcctc ttccagcagg tcaggtgcct gtaacattac aacctcaaaa gcaggttaaa    600
gaaaataaga cccaaccgcc agtagcctat caatactggc ctccggctga acttcagtat   660
cggccacccc cagaaagtca gtatggatat ccaggaatgc ccccagcacc acagggcagg   720
gcgccatacc ctcagccgcc cactaggaga cttaatccta cggcaccacc tagtagacag   780
ggtagtaaat tacatgaaat tattgataaa tcaagaaagg aaggagatac tgaggcatgg   840
caattcccag taacgttaga accgatgcca cctggagaag gagcccaga .gggagagcct   900
cccacagttg aggccagata caagtctttt tcgataaaaa agctaaaaga tatgaaagag   960
ggagtaaaac agtatggacc caactcccct tatatgagga cattattaga ttccattgct   1020
catggacata gactcattcc ttatgattgg gagattctgg caaaatcgtc tctctcaccc   1080
tctcaatttt tacaatttaa gacttggtgg attgatgggg tacaagaaca ggtccgaaga   1140
aataggggctg ccaatcctcc agttaacata gatgcagatc aactattagg aataggtcaa   1200
aattggagta ctattagtca acaagcatta atgcaaaatg aggccattga gcaagttaga   1260
gctatctgcc ttagagcctg ggaaaaaatc caagacccag gaagtacctg cccctcattt   1320
aatacagtaa gacaaggttc aaaagagccc tatcctgatt ttgtggcaag gctccaagat   1380
gttgctcaaa agtcaattgc tgatgaaaaa gcccgtaagg tcatagtgga gttgatggca   1440
tatgaaaacg ccaatcctga gtgtcaatca gccattaagc cattaaaagg aaaggttcct   1500
gcaggatcag atgtaatctc agaatatgta aaagcctgtg atggaatcgg aggagctatg   1560
cataaagcta tgcttatggc tcaagcaata acaggagttg ttttaggagg acaagttaga   1620

Page 60

```
acatttggaa gaaaatgtta taattgtggt caaattggtc acttaaaaaa gaattgccca   1680
gtcttaaata aacagaatat aactattcaa gcaactacaa caggtagaga gccacctgac   1740
ttatgtccaa gatgtaaaaa aggaaaacat tgggctagtc aatgtcgttc taaatttgat   1800
aaaaatgggc aaccattgtc gggaaacgag caaaggggcc agcctcaggc cccacaacaa   1860
actggggcat tcccaattca gccatttgtt cctcagggtt ttcagggaca acaacccca    1920
ctgtcccaag tgtttcaggg aataagccag ttaccacaat acaacaattg tcccccgcca   1980
caagcggcag tgcagcagta g                                            2001
```

<210> 62
<211> 2004
<212> DNA
<213> Artificial Sequence

<220>
<223>   Modified gag sequence

<400>   62
```
atgggccaga ccaagagcaa gatcaagagc aagtacgcca gctacctgag cttcatcaag   60
atcctgctga agcgcggcgg cgtgaaggtg agcaccaaga acctgatcaa gctgttccag   120
atcatcgagc agttctgccc ctggttcccc gagcagggca ccctggacct gaaggactgg   180
aagcgcatcg gcaaggagct gaagcaggcc ggccgcaagg gcaacatcat ccccctgacc   240
gtgtggaacg actgggccat catcaaggcc gccctggagc ccttccagac cgaggaggac   300
agcgtgagcg tgagcgacgc ccccggcagc tgcatcatcg actgcaacga gaacacccgc   360
aagaagagcc agaaggagac cgagggcctg cactgcgagt acgtggccga gcccgtgatg   420
gcccagagca cccagaacgt ggactacaac cagctgcagg aggtgatcta ccccgagacc   480
ctgaagctgg agggcaaggg ccccgagctg gtgggcccca gcgagagcaa gccccgcggc   540
accagccccc tgcccgccgg ccaggtgccc gtgaccctgc agccccagaa gcaggtgaag   600
gagaacaaga cccagccccc cgtggcctac cagtactggc ccccgccga gctgcagtac   660
cgcccccccc ccgagagcca gtacggctac cccggcatgc cccccgcccc ccagggccgc   720
gcccctacc cccagccccc cacccgccgc ctgaacccca ccgccccccc cagccgccag   780
ggcagcaagc tgcacgagat catcgacaag agccgcaagg agggcgacac cgaggcctgg   840
cagttccccg tgaccctgga gcccatgccc cccggcgagg gcgcccagga gggcgagccc   900
cccaccgtgg aggcccgcta caagagcttc agcatcaaga agctgaagga catgaaggag   960
ggcgtgaagc agtacggccc caacagcccc tacatgcgca ccctgctgga cagcatcgcc   1020
cacggccacc gcctgatccc ctacgactgg gagatcctgg ccaagagcag cctgagcccc   1080
agccagttcc tgcagttcaa gacctggtgg atcgacggcg tgcaggagca ggtgcgccgc   1140
aaccgcgccg ccaaccccc cgtgaacatc gacgccgacc agctgctggg catcggccag   1200
aactggagca ccatcagcca gcaggccctg atgcagaacg aggccatcga gcaggtgcgc   1260
gccatctgcc tgcgcgcctg ggagaagatc caggaccccg gcagcacctg ccccagcttc   1320
aacaccgtgc gccagggcag caaggagccc tacccccgact tcgtggcccg cctgcaggac   1380
gtggcccaga gagcatcgc cgacgagaag gcccgcaagg tgatcgtgga gctgatggcc   1440
tacgagaacg ccaaccccga gtgccagagc gccatcaagc ccctgaaggg caaggtgccc   1500
gccggcagcg acgtgatcag cgagtacgtg aaggcctgcg acggcatcgg cggcgccatg   1560
cacaaggcca tgctgatggc ccaggccatc accggcgtgg tgctgggcgg ccaggtgcgc   1620
accttcggcc gcaagtgcta caactgcggc cagatcggcc acctgaagaa gaactgcccc   1680
gtgctgaaca agcagaacat caccatccag gccaccacca ccggccgcga cccccccgac   1740
ctgtgccccc gctgcaagaa gggcaagcac tgggccagcc agtgccgcag caagttcgac   1800
aagaacggcc agcccctgag cggcaacgag cagcgcggcc agccccaggc ccccagcag   1860
accggcgcct tccccatcca gcccttcgtg ccccagggct ccagggcca gcagccccc    1920
ctgagccagg tgttccaggg catcagccag ctgccccagt acaacaactg cccccccccc   1980
caggccgccg tgcagcaggc ttaa                                         2004
```

<210> 63
<211> 1005
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400>   63
```
atgtggggcaa ccattgtcgg gaaacgagca aaggggccag cctcaggccc cacaacaaac  60
tggggcattc ccaattcagc catttgttcc tcagggtttt cagggacaac aaccccccact  120
gtcccaagtg tttcagggaa taagccagtt accacaatac aacaattgtc ccccgccaca  180
agcggcagtg cagcagtaga tttatgtact atacaagcag tctctctgct tccaggggag  240
ccccccacaaa aaacccccac aggggtatat ggacccctgc ctaaggggac tgtaggacta  300
```

```
atcttgggac gatcaagtct aaatctaaaa ggagttcaaa ttcatactag tgtggttgat    360
tcagactata aaggcgaaat tcaattggtt attagctctt caattccttg gagtgccagt    420
ccaagagaca ggattgctca attattactc ctgccataca ttaagggtgg aaatagtgaa    480
ataaaaagaa taggagggct tggaagcact gatccaacag gaaaggctgc atattgggca    540
agtcaggtct cagagaacag acctgtgtgt aaggccatta ttcaaggaaa acagtttgaa    600
gggttggtag acactggagc agatgtctct atcattgctt taaatcagtg gccaaaaaat    660
tggcctaaac aaaaggctgt tacaggactt gtcggcatag gcacagcctc agaagtgtat    720
caaagtacgg agattttaca ttgcttaggg ccagataatc aagaaagtac tgttcagcca    780
atgattactt caattcctct taatctgtgg ggtcgagatt tattacaaca atggggtgcg    840
gaaatcacca tgcccgctcc atcatatagc cccacgagtc aaaaaatcat gaccaagatg    900
ggatatatac caggaaaggg actagggaaa aatgaagatg gcattaaaat tccagttgag    960
gctaaaataa atcaagaaag agaaggaata gggaatcctt gctag                   1005
```

```
<210>   64
<211>   1008
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Modified Prt sequence

<400>   64
atgtgggcca ccatcgtggg caagcgcgcc aagggccccg ccagcggccc caccaccaac    60
tggggcatcc ccaacagcgc catctgcagc agcggcttca gcggcaccac cacccccacc    120
gtgcccagcg tgagcggcaa caagcccgtg accaccatcc agcagctgag ccccgccacc    180
agcggcagcg ccgccgtgga cctgtgcacc atccaggccg tgagcctgct gcccggcgag    240
cccccccaga agacccccac cggcgtgtac ggccccctgc ccaagggcac cgtgggcctg    300
atcctgggcc gcagcagcct gaacctgaag ggcgtgcaga tccacaccag cgtggtggac    360
agcgactaca agggcgagat ccagctggtg atcagcagca gcatccccctg gagcgccagc    420
ccccgcgacc gcatcgccca gctgctgctg ctgccctaca tcaagggcgg caacagcgag    480
atcaagcgca tcggcggcct gggcagcacc gaccccaccg gcaaggccgc ctactgggcc    540
agccaggtga gcgagaaccg ccccgtgtgc aaggccatca tccagggcaa gcagttcgag    600
ggcctggtgg acaccggcgc cgacgtgagc atcatcgccc tgaaccagtg gcccaagaac    660
tggcccaagc agaaggccgt gaccggcctg gtgggcatcg gcaccgccag cgaggtgtac    720
cagagcaccg agatcctgca ctgcctgggc cccgacaacc aggagagcac cgtgcagccc    780
atgatcacca gcatccccct gaacctgtgg ggccgcgacc tgctgcagca gtggggcgcc    840
gagatcacca tgcccgcccc cagctacagc cccaccagcc agaagatcat gaccaagatg    900
ggctacatcc ccggcaaggg cctgggcaag aacgaggacg gcatcaagat ccccgtggag    960
gccaagatca accaggagcg cgagggcatc ggcaacccct gcgcttaa                 1008·
```

```
<210>   65
<211>   2874
<212>   DNA
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   65
atgaataaat caagaaagag aaggaatagg gaatccttgc taggggcggc cactgtagag    60
cctcctaaac ccataccatt aacttggaaa acagaaaaac cagtgtgggt aaatcagtgg    120
ccgctaccaa aacaaaaact ggaggcttta catttattag caaatgaaca gttagaaaag    180
ggtcatattg agccttcgtt ctcaccttgg aattctcctg tgtttgtaat tcagaagaaa    240
tcaggcaaat ggcgtatgtt aactgactta agggctgtaa acgccgtaat tcaacccatg    300
gggcctctcc aacccgggtt gccctctccg gccatgatcc caaaagattg cccttttaatt    360
ataattgatc taaaggattg cttttttacc atccctctgg cagagcagga ttgcgaaaaa    420
tttgccttta ctataccagc cataaataat aaagaaccag ccaccaggtt tcagtggaaa    480
gtgttacctc agggaatgct taatagtcca actatttgtc agactttgt aggtcgagct    540
cttcaaccag ttagagaaaa gttttcagac tgttatatta ttcattgtat tgatgatatt    600
ttatgtgctg cagaaacgaa agataaatta attgactgtt atacatttct gcaagcagag    660
gttgccaatg ctggactggc aatagcatct gataagatcc aaacctctac tccttttcat    720
tatttaggga tgcagataga aaatagaaaa attaagccac aaaaaataga aataagaaaa    780
gacacattaa aaacactaaa tgattttcaa aaattactag gagatattaa ttggattcgg    840
ccaactctag gcattcctac ttatgccatg tcaaatttgt tctctatctt aagaggagac    900
tcagacttaa atagtaaaag aatgttaacc ccagaggcaa caaagaaat taaattagtg    960
gaagaaaaaa ttcagtcagc gcaaataaat agaatagatc ccttagcccc actccaactt    1020
```

```
ttgatttttg ccactgcaca ttctccaaca ggcatcatta ttcaaaatac tgatcttgtg   1080
gagtggtcat tccttcctca cagtacagtt aagactttta cattgtactt ggatcaaata   1140
gctacattaa tcggtcagac aagattacga ataataaaat tatgtgggaa tgacccagac   1200
aaaatagttg tccctttaac caaggaacaa gttagacaag cctttatcaa ttctggtgca   1260
tggaagattg gtcttgctaa ttttgtggga attattgata atcattaccc aaaaacaaag   1320
atcttccagt tcttaaaatt gactacttgg attctaccta aaattaccag acgtgaacct   1380
ttagaaaatg ctctaacagt atttactgat ggttccagca atggaaaagc agcttacaca   1440
ggaccgaaag aacgagtaat caaaactcca tatcaatcgg ctcaaagagc agagttggtt   1500
gcagtcatta cagtgttaca agattttgac caacctatca atattatatc agattctgca   1560
tatgtagtac aggctacaag ggatgttgag acagctctaa ttaaatatag catggatgat   1620
cagttaaacc agctattcaa tttattacaa caaactgtaa gaaaaagaaa tttcccattt   1680
tatattacac atattcgagc acacactaat ttaccagggc ctttgactaa agcaaatgaa   1740
caagctgact tactggtatc atctgcactc ataaaagcac aagaacttca tgctttgact   1800
catgtaaatg cagcaggatt aaaaaacaaa tttgatgtca catggaaaca ggcaaaagat   1860
attgtacaac attgcaccca gtgtcaagtc ttacacctgc ccactcaaga ggcaggagtt   1920
aatcccagag gtctgtgtcc taatgcatta tggcaaatgg atgtcacgca tgtaccttca   1980
tttggaagat tatcatatgt tcacgtaaca gttgatactt attcacattt catatgggca   2040
acttgccaaa caggagaaag tacttcccat gttaaaaaac atttattgtc ttgttttgct   2100
gtaatgggag ttccagaaaa aatcaaaact gacaatggac caggatattg tagtaaagct   2160
ttccaaaaat tcttaagtca gtggaaaatt tcacatacaa caggaattcc ttataattcc   2220
caaggacagg ccatagttga aagaactaat agaacactca aaactcaatt agttaaacaa   2280
aaagaagggg gagacagtaa ggagtgtacc actcctcaga tgcaacttaa tctagcactc   2340
tatactttaa attttttaaa catttataga aatcagacta ctacttctgc agaacaacat   2400
cttactggta aaaagaacag cccacatgaa ggaaaactaa tttggtggaa agataataaa   2460
aataagacat gggaaatagg gaaggtgata acgtggggga gaggttttgc ttgtgtttca   2520
ccaggagaaa atcagcttcc tgtttggata cccactagac atttgaagtt ctacaatgaa   2580
cccatcagag atgcaaagaa aagcacctcc gcggagacgg agacatcgca atcgagcacc   2640
gttgactcac aagatgaaca aaatggtgac gtcagaagaa cagatgaagt tgccatccac   2700
caagaaggca gagccgccaa cttgggcaca actaaagaag ctgacgcagt tagctacaaa   2760
atatctagag aacacaaagg tgacacaaac cccagagagt atgctgcttg cagccttgat   2820
gattgtatca atggtggtaa gtctccctat gcctgcagga gcagctgcag ctaa          2874
```

<210> 66
<211> 2877
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified pol sequence

<400> 66
```
atgaacaaga gccgcaagcg ccgcaaccgc gagagcctgc tgggcgccgc caccgtggag    60
cccccaagc ccatcccct gacctggaag accgagaagc ccgtgtgggt gaaccagtgg    120
cccctgccca agcagaagct ggaggccctg cacctgctgg ccaacgagca gctggagaag   180
ggccacatcg agcccagctt cagcccctgg aacagccccg tgttcgtgat ccagaagaag   240
agcggcaagt ggcgcatgct gaccgacctg cgcgccgtga cgccgtgat ccagcccatg    300
ggccccctgc agcccggcct gcccagcccc gccatgatcc caaggactg gcccctgatc    360
atcatcgacc tgaaggactg cttcttcacc atcccctgg ccgagcagga ctgcgagaag    420
ttcgccttca ccatccccgc catcaacaac aaggagcccg ccacccgctt ccagtggaag   480
gtgctgcccc agggcatgct gaacagcccc accatctgcc agaccttcgt gggccgcgcc   540
ctgcagcccg tgcgcgagaa gttcagcgac tgctacatca tccactgcat cgacgacatc   600
ctgtgcgccg ccgagaccaa ggacaagctg atcgactgct acaccttcct gcaggccgag   660
gtggccaacg ccggcctggc catcgccagc gacaagatcc agaccagcac ccccttccac   720
tacctgggca tgcagatcga gaaccgcaag atcaagcccc agaagatcga gatccgcaag   780
gacaccctga gaccctgaa cgacttccag aagctgctgg cgacatcaa ctggatccgc    840
cccaccctgg gcatccccac ctacgccatg agcaacctgt cagcatcct gcgcggcgac    900
agcgacctga acagcaagcg catgctgacc cccgaggcca ccaaggagat caagctggtg   960
gaggagaaga tccagagcgc ccagatcaac cgcatcgacc ccctggcccc cctgcagctg   1020
ctgatcttcg ccaccgccca gcccccacc ggcatcatca tccagaacac cgacctggtg   1080
gagtggagct tcctgcccca gcaccgtg aagaccttca ccctgtacct ggaccagatc    1140
gccaccctga tcggccagac ccgcctgcgc atcatcaagc tgtgcggcaa cgaccccgac   1200
aagatcgtgg tgcccctgac caaggagcag gtgcgccagg ccttcatcaa cagcggcgcc   1260
tggaagatcg gctggccaa cttcgtgggc atcatcgaca ccactaccc caagaccaag   1320
```

```
atcttccagt tcctgaagct gaccacctgg atcctgccca agatcacccg ccgcgagccc   1380
ctggagaacg ccctgaccgt gttcaccgac ggcagcagca acggcaaggc cgcctacacc   1440
ggccccaagg agcgcgtgat caagacccc taccagagcg cccagcgcgc cgagctggtg   1500
gccgtgatca ccgtgctgca ggacttcgac cagcccatca acatcatcag cgacagcgcc   1560
tacgtggtgc aggccacccg cgacgtggag accgccctga tcaagtacag catggacgac   1620
cagctgaacc agctgttcaa cctgctgcag cagaccgtgc gcaagcgcaa cttcccccttc   1680
tacatcaccc acatccgcgc ccacaccaac ctgcccggcc ccctgaccaa ggccaacgag   1740
caggccgacc tgctggtgag cagcgccctg atcaaggccc aggagctgca cgccctgacc   1800
cacgtgaacg ccgccggcct gaagaacaag ttcgacgtga cctggaagca ggccaaggac   1860
atcgtgcagc actgcaccca gtgccaggtg ctgcacctgc cacccagga ggccggcgtg   1920
aacccccgcg gcctgtgccc caacgccctg tggcagatgg acgtgaccca cgtgcccagc   1980
ttcggccgcc tgagctacgt gcacgtgacc gtggacacct acagccactt catctgggcc   2040
acctgccaga ccggcgagag caccagccac gtgaagaagc acctgctgag ctgcttcgcc   2100
gtgatgggcg tgcccgagaa gatcaagacc gacaacggcc ccggctactg cagcaaggcc   2160
ttccagaagt tcctgagcca gtggaagatc agccacacca ccggcatccc ctacaacagc   2220
cagggccagg ccatcgtgga gcgcaccaac cgcaccctga gacccagct ggtgaagcag   2280
aaggagggcg gcgacagcaa ggagtgcacc acccccagga tgcagctgaa cctggcccctg   2340
tacaccctga acttcctgaa catctaccgc aaccagacca ccaccagcgc cgagcagcac   2400
ctgaccggca gaagaacag ccccacgag ggcaagctga tctggtggaa ggacaacaag   2460
aacaagacct gggagatcgg caaggtgatc acctgggggcc gcggcttcgc ctgcgtgagc   2520
cccggcgaga accagctgcc cgtgtggatc cccaccccgcc acctgaagtt ctacaacgag   2580
cccatccgcg acgccaagaa gagcgaccagc gccgagaccg agaccagcca gagcagcacc   2640
gtggacagcc aggacgagca gaacggcgac gtgcgccgca ccgacgaggt ggccatccac   2700
caggagggcc gcgccgccaa cctgggcacc accaaggagg ccgacgccgt gagctacaag   2760
atcagccgcg agcacaaggg cgacaccaac ccccgcgagt acgccgcctg cagcctggac   2820
gactgcatca acggcggcaa gagcccctac gcctgccgca gcagctgcag cgcttaa     2877
```

<210> 67
<211> 106
<212> PRT
<213> Artificial Sequence

<220>
<223> Manipulated cORF

<400> 67

```
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
                85                  90                  95

Glu Thr Ala Thr Ile Glu Asn Gly Pro Ala
            100                 105
```

<210> 68
<211> 145
<212> PRT
<213> Artificial Sequence

<220>

<223> Manipulated PCAP5

<400> 68
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
            85                  90                  95

Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys
            100                 105                 110

Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
            115                 120                 125

Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
    130                 135                 140

Ala
145

<210>   69
<211>   666
<212>   PRT
<213>   Human endogenous retrovirus, K family (HERV-K)

<400>   69
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1               5                   10                  15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
            20                  25                  30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
        35                  40                  45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
    50                  55                  60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65                  70                  75                  80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
            85                  90                  95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
            100                 105                 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
            115                 120                 125

```
Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
    130             135             140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145             150             155             160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
                165             170             175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
            180             185             190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
            195             200             205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
    210             215             220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225             230             235             240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
            245             250             255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
            260             265             270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
        275             280             285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
    290             295             300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305             310             315             320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
            325             330             335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
            340             345             350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
            355             360             365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
    370             375             380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385             390             395             400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
            405             410             415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
            420             425             430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
            435             440             445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
    450             455             460
```

Page 66

```
Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465             470             475             Leu     480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
                485             490             .       495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
            500             505         .       510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
            515             520             525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
    530             535             540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545             550             555             560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
                565             570             575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
            580             585             590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
        595             600             605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
    610             615             620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625             630             635             640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
                645             650             655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
            660             665
```

```
<210>   70
<211>   667
<212>   PRT
<213>   Artificial Sequence

<220>
<223>   Manipulated Gag

<400>   70
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1               5               10              15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
            20              25              30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
        35              40              .       45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
    50              55              60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65              70              75              80
```

Page 67

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
                85                  90                  95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
            100             105             110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
        115             120             125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
    130             135             140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145             150             155             160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
            165             170             175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
            180             185             190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
            195             200             205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
    210             215             220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225             230             235             240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
            245             250             255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
            260             265             270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
            275             280             285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
    290             295             300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305             310             315             320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
            325             330             335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
            340             345             350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
            355             360             365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
    370             375             380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385             390             395             400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
            405             410             415

Page 68

```
Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
            420             425             430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
            435             440             445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
    450             455             460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465             470             475             480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
            485             490             495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
            500             505             510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
            515             520             525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
    530             535             540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545             550             555             560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
            565             570             575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
            580             585             590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
            595             600             605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
    610             615             620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625             630             635             640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
            645             650             655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln Ala
            660             665
```

```
<210>    71
<211>    334
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    71
Met Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly
1               5               10              15

Pro Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly
            20              25              30

Phe Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys
    35              40              45
```

```
Pro Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala
    50              55              60

Ala Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu
65              70              75              80

Pro Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly
            85              90              95

Thr Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val
            100             105             110

Gln Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln
            115             120             125

Leu Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg
    130             135             140

Ile Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu
145             150             155             160

Ile Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala
            165             170             175

Ala Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala
            180             185             190

Ile Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp
            195             200             205

Val Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln
    210             215             220

Lys Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr
225             230             235             240

Gln Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser
            245             250             255

Thr Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg
            260             265             270

Asp Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser
    275             280             285

Tyr Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro
    290             295             300

Gly Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu
305             310             315             320

Ala Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys
            325             330
```

```
<210>    72
<211>    335
<212>    PRT
<213>    Artificial Sequence

<220>
<223>    Manipulated Prt

<400>    72
```

Page 70

```
Met Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly
1               5                   10                  15

Pro Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly
            20              25              30

Phe Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys
        35              40              45

Pro Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala
    50              55              60

Ala Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu
65              70              75                  80

Pro Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly
            85              90                  95

Thr Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val
            100             105             110

Gln Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln
        115             120             125

Leu Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg
    130             135             140

Ile Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu
145             150             155             160

Ile Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala
            165             170             175

Ala Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala
        180             185             190

Ile Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp
        195             200             205

Val Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln
    210             215             220

Lys Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr
225             230             235             240

Gln Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser
            245             250             255

Thr Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg
            260             265             270

Asp Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser
    275             280             285

Tyr Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro
    290             295             300

Gly Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu
305             310             315             320

Ala Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys Ala
            325             330             335
```

<210> 73
<211> 957
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 73
Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
1               5                   10                  15

Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
            20                  25                  30

Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
            35                  40                  45

Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
        50                  55                  60

Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
65                  70                  75                  80

Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
            85                  90                  95

Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
            100                 105                 110

Ile Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe
            115                 120                 125

Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
        130                 135                 140

Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
145                 150                 155                 160

Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
                165                 170                 175

Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
            180                 185                 190

Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
            195                 200                 205

Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
        210                 215                 220

Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
225                 230                 235                 240

Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
                245                 250                 255

Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
            260                 265                 270

Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
            275                 280                 285

Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
        290                 295                 300

Page 72

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
305                 310                 315                 320

Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
            325                 330                 335

Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
            340                 345                 350

Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
            355                 360                 365

Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
    370                 375                 380

Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
385                 390                 395                 400

Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
            405                 410                 415

Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
            420                 425                 430

Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
            435                 440                 445

Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
    450                 455                 460

Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
465                 470                 475                 480

Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
            485                 490                 495

Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
            500                 505                 510

Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
            515                 520                 525

Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
    530                 535                 540

Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
545                 550                 555                 560

Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
            565                 570                 575

Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
            580                 585                 590

Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
            595                 600                 605

Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
    610                 615                 620

Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
625                 630                 635                 640

Asn Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr
645           650            655

His Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp
660           665         670

Thr Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr
675          680          685

Ser His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val
690          695         700

Pro Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala
705          710        715       720

Phe Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile
725          730        735

Pro Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr
740          745        750

Leu Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu
755          760        765

Cys Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn
770          775        780

Phe Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His
785          790        795      800

Leu Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp
805          810        815

Lys Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp
820          825        830

Gly Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val
835          840        845

Trp Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp
850          855        860

Ala Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr
865          870        875      880

Val Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu
885          890        895

Val Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys
900          905        910

Glu Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp
915          920        925

Thr Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn
930          935        940

Gly Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser
945          950        955

&lt;210&gt;   74
&lt;211&gt;   958
&lt;212&gt;   PRT

&lt;213&gt;    Artificial Sequence

&lt;220&gt;
&lt;223&gt;    Manipulated Pol

&lt;400&gt;    74
```
Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
1               5                   10                  15

Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
                20                  25                  30

Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
            35                  40                  45

Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
        50                  55                  60

Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
65                  70                  75                  80

Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
                85                  90                  95

Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
            100                 105                 110

Ile Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe
        115                 120                 125

Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
    130                 135                 140

Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
145                 150                 155                 160

Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
                165                 170                 175

Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
            180                 185                 190

Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
        195                 200                 205

Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
    210                 215                 220

Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
225                 230                 235                 240

Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
                245                 250                 255

Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
            260                 265                 270

Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
        275                 280                 285

Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
    290                 295                 300
```

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
305                310                315                320

Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
                 325            330                335

Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
             340                345                350

Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
         355                360                365

Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
    370                375                380

Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
385                390                395                400

Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
             405                410                415

Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
             420                425                430

Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
         435                440                445

Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
    450                455                460

Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
465                470                475                480

Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
             485                490                495

Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
             500                505                510

Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
         515                520                525

Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
    530                535                540

Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
545                550                555                560

Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
             565                570                575

Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
             580                585                590

Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
             595                600                605

Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
    610                615                620

Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
625                630                635                640

**Page 76**

Asn Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr
                645             650                 655

His Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp
            660             665             670

Thr Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr
        675             680             685

Ser His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val
    690             695             700

Pro Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala
705             710             715             720

Phe Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile
            725             730             735

Pro Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr
            740             745             750

Leu Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu
        755             760             765

Cys Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn
    770             775             780

Phe Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His
785             790             795             800

Leu Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp
            805             810             815

Lys Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp
            820             825             830

Gly Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val
        835             840             845

Trp Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp
    850             855             860

Ala Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr
865             870             875             880

Val Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu
            885             890             895

Val Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys
            900             905             910

Glu Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp
        915             920             925

Thr Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn
    930             935             940

Gly Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser Ala
945             950             955

<210>    75
<211>    12366
<212>    DNA

Page 77

<213>    Human endogenous retrovirus, K family (HERV-K), located at 22q11.2

<400>    75
tgtggggaaa agaaagagag atcagactgt tactgtgtct atgtagaaag aaatagacat    60
aagagactcc attttgttct gtactaagaa aaattcttct gctttgagat gctgttaatc    120
tgtaacccta gccccaaccc tgtgctcaca gaaacaggtg ctgtgttgac tcaaggttta    180
atggattcag ggctgtgcag gatgtgcttt gttaaacaaa tgcttgaagg cagcaagctt    240
gttaagagtc atcaccactc cctaatctca agtaagcagg gacacaaaca ctgcggaagg    300
ccgcagggac ctctgcctag gaaagccagg tgttgtccaa ggtttctccc catgtgacag    360
tctgaaatat ggcctcttgg gaagggaaag acctgactgt cccctggccc gacacccgta    420
aagggtctgt gctgaggatt agtaaaagag gaaggaaggc ctctttgcag ttgagataag    480
aggaaggcat ctgtctcctg ctcatccctg ggcaatggaa tgtcttggtg taaagcctga    540
ttgtatatgc catctactga gataggagaa aactgcctta gggctggagg tgggacatgc    600
tggcggcaat actgctcttt aaggcattga gatgtttatg tatatgcaca tcaaaagcac    660
agcacttttt tctttacctt gtttatgatg cagagacatt tgttcacatg ttttcctgct    720
ggccctctcc ccactattac cctattgtcc tgccacatcc ccctctccga gatggtagag    780
ataatgatca ataaatactg agggaactca gagaccggtg cggcgcgggt cctccatatg    840
ctgagcgccg gtcccctggg cccacttttc tttctctata ctttgtctct gttgtctttc    900
ttttctcaag tctctcgttc cacctgagga gaaatgccca cagctgtgga ggcgcaggcc    960
actccatctg gtgcccaacg tggatgcttt tctctagggt gaagggactc tcgagtgtgg    1020
tcattgagga caagtcaacg agagattccc gagtacgtct acagtgagcc ttgtggtaag    1080
cttgggcgct cggaagaagc caggggttaat ggggcaaact aaaagtaaag tctctcattc    1140
cacctgatga gaaacaccca gaggtgtgga ggggcaggcc accccttcag ggtagggtcc    1200
cctccatgca gaccatagag cacaggtgtg ccccaaagag gagcagagag aaggagggag    1260
agggcccacg agagacttgg aaatgaatgg caggatttta ggcgctggac ttgggttcgg    1320
ggcacctggc cttttccttgt gtatttctcc tactgtctgc ctaactattt aatacaataa    1380
aagaaaacca gcccctggtt cttgtggtgt ttccaccctc ccgggtcccc gctggctgcc    1440
tggcttcctc ccgcagctcc tgctgtgtgt gtatgtgtgt gtgtgtgcac atctgtgggg    1500
cgtatgtgtg ttcgtctttg taattgaggc tgcagagtgg agagagcagg ggttttctct    1560
ggggacccag agagaaggag gcgtttttcac cacagccgaa cagggcagga ccccagcacc    1620
cgggacccag cgggactttg ccaaggggat ggacctggct gggccacgcg gctgtttgtg    1680
tagggaaaag aaagagagat cacactgtta ctgtgtctat gtagaaaagg aagacataaa    1740
ctccattttg agctgtacta agaaaaatta ttttgccttg acctgctgtt aacctgtaac    1800
tgtagcccca accctgtgct caaagaaaca tgtgctgtat ggaatcaagg tttaagggat    1860
caagggctgt acaggatgtg ccttgttaac aatgtgttta caggcagtat gcttggtaaa    1920
agtcatcgcc attctccatt ctccattaat caggggcacg atgcactgcg gaaagccaca    1980
gggacctctg cccgagaaag cctgggtatt gtccaaggct tcccccccact gagacagcct    2040
gagatacggc ctcgtgggaa gggaaagacc tgaccgtccc ccagcccgac acccgtaaag    2100
ggtctgtgct gaggaggatt agtaaaaggg gaaggcctct tgcagttgag ataagaggaa    2160
ggcctccgtc tcctgcatgt ccttgggaat ggaatgtctt ggtgtaaaac ccgatagtac    2220
attccttcta ttctgagaga agaaaccac cctgtggctg gaggtgagat atgctagcgg    2280
caatgctgct ctgttactct ttgctacact gagatgtttg ggtggagaga agcataaatc    2340
tggcctatgt gcacatctgg gcacagaacc tcccccttgaa cttgtgacac agattccttt    2400
gttcacatgt tttcctgctg accttctccc cactatcgcc ctgttctccc accgcattcc    2460
ccttgctgag atagtgaaaa tagtaatctg tagataccaa gggaactcag agaccatggc    2520
cggtgcacat cctccgtacg ctgagcgctg gtcccctggg cccattgttc tttctctata    2580
ctttgtctct gtgtcttatt tctttcctca gtctctcatc cctcctgacg agaaataccc    2640
acaggtgtgg aggggctggc ccccttcatc tgatgcccaa tgtgggtgcc tttctctagg    2700
gtgaaggtac tctacagtgt ggtcattgag gacaagttga cgagagagtc ccaagtacgt    2760
ccacggtcag ccttgcggta agcttgtgtg cttagaggaa cccagggtaa cgatggggca    2820
aactgaaagt aaatatgcct cttatctcag ctttattaaa attctttaa gaagaggggg    2880
agttagagct tctacagaaa atctaattac gctatttcaa acaatagaac aattctgccc    2940
atggtttcca gaacagggaa ctttagatct aaaagattgg gaaaaaattg gcaaagaatt    3000
aaaacaagca aatagggaag gtaaaatcat cccacttaca gtatggaatg attgggccat    3060
tattaaagca actttagaac catttcaaac aggagaagat attgtttcag tttctgatgc    3120
ccctaaaagc tgtgtaacag attgtgaaga agaggcaggg acagaatccc agcaaggaac    3180
ggaaagttca cattgtaaat atgtagcaga gtctgtaatg gctcagtcaa cgcaaaatgt    3240
tgactacagt caattacagg agataatata ccctgaatca tcaaaattgg gggaaggagg    3300
tccagaatca ttggggccat cagagcctaa accacgatcg ccatcaactc ctcctcccgt    3360
ggttcagatg cctgtaacat tacaacctca aacgcaggtt agacaagcac aaaccccaag    3420
agaaaatcaa gtagaaaggg acagagtctc tatcccggca atgccaactc agatacagta    3480
tccacaatat cagccggtag aaaataagac ccaaccgctg gtagtttatc aataccggct    3540
gccaaccgag cttcagtatc ggcctccttc agaggttcaa tacagacctc aagcggtgtg    3600

```
tcctgtgcca aatagcacgg caccatacca gcaacccaca gcgatggcgt ctaattcacc   3660
agcaacacag gacgcggcgc tgtatcctca gccgcccact gtgagactta atcctacagc   3720
atcacgtagt ggacagggtg gtgcactgca tgcagtcatt gatgaagcca gaaaacaggg   3780
cgatcttgag gcatggcggt tcctggtaat tttacaactg gtacaggccg gggaagagac   3840
tcaagtagga gcgcctgccc gagctgagac tagatgtgaa cctttcacca tgaaaatgtt   3900
aaaagatata aaggaaggag ttaaacaata tggatccaac tcccttata taagaacatt   3960
attagattcc attgctcatg gaaatagact tactccttat gactgggaaa ttttggccaa   4020
atcttccctt tcatcctctc agtatctaca gtttaaaacc tggtggattg atggagtaca   4080
agaacaggta cgaaaaaatc aggctactaa gcccactgtt aatatagacg cagaccaatt   4140
gttaggaaca ggtccaaatt ggagcaccat taaccaacaa tcagtgatgc agaatgaggc   4200
tattgaacaa gtaagggcta tttgcctcag ggcctgggga aaaattcagg acccaggaac   4260
agctttccct attaattcaa ttagacaagg ctctaaagag ccatatcctg actttgtggc   4320
aagattacaa gatgctgctc aaaagtctat tacagatgac aatgcccgaa aagttattgt   4380
agaattaatg gcctatgaaa atgcaaatcc agaatgtcag tcggccataa agccattaaa   4440
aggaaaagtt ccagcaggag ttgatgtaat tacagaatat gtgaaggctt gtgatgggat   4500
tggaggagct atgcataagg caatgctaat ggctcaagca atgaggggc tcactctagg   4560
aggacaagtt agaacatttg ggaaaaaatg ttataattgt ggtcaaatcg gtcatctgaa   4620
aaggagttgc ccaggcttaa ataaacagaa tataataaat caagctatta cagcaaaaaa   4680
taaaaagcca tctggcctgt gtccaaaatg tggaaaagca aaacattggg ccaatcaatg   4740
tcattctaaa tttgataaag atgggcaacc attgtctgga aacaggaaga ggggccagcc   4800
tcaggccccc caacaaactg gggcattccc agttaaactg tttgttcctc agggttttca   4860
aggacaacaa cccctacaga aaataccacc acttcaggga gtcagccaat tacaacaatc   4920
caacagctgt cccgcgccac agcaggcagc accgcagtag atttatgttc cacccaaatg   4980
gtctttttac tccctggaaa gccccacaa aagattccta gaggggtata tggcccgctg   5040
ccagaaggga gggtaggcct ttgagggaga tcaagtctaa atttgaaggg agtccaaatt   5100
catactgggg taatttattc agattataaa gggggaattc agttagtgat cagctccact   5160
gttccccgga gtgccaatcc aggtgataga attgctcaat tactgctttt gccttatgtt   5220
aaaattgggg aaaacaaaaa ggaaagaaca ggagggtttg gaagtaccaa ccctgcagga   5280
aaagctgctt attgggctaa tcaggtctca gaggatagac ccgtgtgtac agtcactatt   5340
cagggaaaga gtttgaagga ttagtggata cccaggctga tgtttctgtc atcggcatag   5400
gtactgcctc agaagtgtat caaagtgcca tgattttaca ttgtccagga tctgataatc   5460
aagaaagtac ggttcagcct gtgatcactt cattccaatc aatttatggg gccgagactt   5520
gttacaacaa tggcatgcag agattactat cccagcctcc ctatacagcc ccaggaataa   5580
aaaaatcatg actaaaatgg gatagctccc taaaaaggga ctaggaaaga agtcccaatt   5640
gaggctgaaa aaaatcaaaa aagaaaagga atagggcatc ctttttagga gcggtcactg   5700
tagagcctcc aaaacccatt ccattaactt gggggaaaaa aaaacaactg tatggtaaat   5760
cagcagcgct tccaaaacaa aaactggagg ctttacattt attagcaaag aaacaattag   5820
aaaaaggaca ttgagccttc attttcgcct tggaattctg tttgtaattc agaaaaaatc   5880
cggcagatgg cgtataatgc cgtaattcaa cccatggggg ctctcccacc ccggttgccc   5940
tctccagcca tggtcccctt taattataat tgatctgaag gattgctttt ttaccattcc   6000
tctggcaaaa caggattttg aaaaatttgc ttttaccaca ccagcctaaa taataaagaa   6060
ccagccacca ggtttcagtg gaaagtattg cctcagggaa tgcttaatag ttcaactatt   6120
tgtcagctca agctctgcaa ccagttagag acaagttttc agactgttac atcgttcact   6180
atgttgatat tttgtgtgct gcagaaacga gagacaaatt aattgaccgt tacacatttc   6240
tgcagacaga ggttgccaac gcgggactga caataacatc tgataagatt caaacctcta   6300
ctcctttccg ttacttggga atgcaggtag aggaaaggaa aattaaacca caaaaaatag   6360
aaataagaaa agacacatta aaagcattaa atgagtttca aaagttgcta ggagatacta   6420
attggatttg gagatattaa ttggatttgg ccaactctag gcattcctac ttatgccatg   6480
tcaaatttgt tctctttctt aagaggggac tcggaattaa atagtgaaag aacgttaact   6540
ccagaggcaa ctaaagaaat taaattaatt gaagaaaaaa ttcggtcagc acaagtaaat   6600
agaatagatc acttggcccc actccaaatt ttgattttg ctactgcaca ttccctaaca   6660
ggcatcattg ttcaaaatac agatcttgtg gagtggtcct tccttcctca cagtacaatt   6720
aagacttta cattgtactt ggatcaaatg gctacattaa ttggtcaggg aagattatga   6780
ataataacat tgtgtggaaa tgacccagat aaaatcactg ttcctttcaa caagcaacag   6840
gttagacaag cctttatcaa ttctggtgca tggcagattg gtcttgccga ttttgtggga   6900
attattgaca atcgttaccc caaaacaaaa atcttccagt ttttaaaatt gactacttgg   6960
attttaccta aagttaccaa acataagcct ttaaaaaatg ctctggcagt gtttactgat   7020
ggttccagca atggaaaagt ggcttacacc gggccaaaag aatgagtcat caaaactcag   7080
tatcacttga ctcaaagagc agagttggtt gccgtcatta cagtgttaac aagattttaa   7140
tcagtctatt aacattgtat cagattctgc atatgtagta caggctacaa aggatattga   7200
gagagcccta atcaaataca ttatggatga tcagttaaac ccgctgttta atttgttaca   7260
acaaaatgta agaaaaagaa atttcccatt ttatattact catattcgag cacacactaa   7320
tttaccaggg cctttaacta aagcaaatga acaagctgac ttgctagtat catctgcatt   7380
```

```
catggaagca caagaacttc atgccttgac tcatgtaaat gcaataggat taaaaaataa   7440
atttgatatc acatggaaac agacaaaaaa tattgtacaa cattgcaccc agtgtcagat   7500
tctacacctg gccactcagg aggcaagagt taatcccaga ggtctatgtc ctaatgtgtt   7560
atggcaaatg gatgtcatgc acgtaccttc atttggaaaa ttgtcatttg tccatgtgac   7620
agttgatact tattcacatt tcatatgggc aacctgccag acaggagaaa gtacttccca   7680
tgttaaaaga catttattat cttgtttttcc tgtcatggga gttccagaaa aagttaaaac   7740
agacaatggg ccaggttact gtagtaaagc agttcaaaaa ttcttaaatc agtggaaaat   7800
tacacataca ataggaattc tctataattc ccaaggacag gccataattg aaagaactaa   7860
tagaacactc aaagctcaat tggttaaaca aaaaaaagga aagacagga gtataacact    7920
ccccagatgc aacttaatct agcactctat actttaaatg ttttaaacat ttatagaaat   7980
cagaccacta cctctgcaga acaacatctt actggtaaaa ggaacagccc acatgaagga   8040
aaactgattt ggtggaaaga taataaaaat aaaacatggg aaatggggaa ggtgataacg   8100
tgggggagag gttttgcttg tgtttcacca ggagaaaatc agcttcctgt ttggataccc   8160
actagacatt taaagttcta caatgaactc actggagatg caaagaaaag tgtggagatg   8220
gagacacccc aatcgactcg ccaggtaaac aaaatggtga tatcagaaga acagaaaaag   8280
ttgccttcca tcaaggaagc agagttgcca atataggcac aattaaagaa gctgacacag   8340
ttagctaaaa aaaaaagcct agagaataca aaggtgacac caactccaga gaatatgctg   8400
cttgcagctc tgatgattgt atcaacggtg gtaagtcttc ccaagtctgc aggagcagct   8460
gcagctaatt atacttactg ggcctatgtg cctttcccac ccttaattcg ggcagttaca   8520
tagatggata atcctattga agtagatgtt aataatagtg catgggtgcc tggccccaca   8580
gatgactgtt gccctgccca acctgaagaa ggaatgatga tgaatatttc cattgggtat   8640
ccttatcctc ctgtttgcct agggaaggca ccaggatgct taatgcctac aacccaaaat   8700
tggttggtag aagtacctac agtcagtgct accagtagat ttacttatca catggtaagt   8760
ggaatgtcac agataaataa tttacaggac ccttcttatc aaagatcatt acaatgtagg   8820
cctaagggga aggcttgccc caaggaaatt cccaaagaat caaaaagccc agaagtctta   8880
gtctgcggag aatgtgtggc tgatactgca gtgtagtaca aaacaatgaa ttttgaacta   8940
tgatagactg ggtcccttga ggccaattat atcataactg tacaggccag actcattcat   9000
gttcacaggc cccatccatc tggcccatta atccagccta tgacggtgat gtaactgaaa   9060
ggctggacca ggtttataga aggttagaat cactctgtcc aaggaaatgg ggtgaaaagg   9120
gaatttcatc accttgacca aagttagtcc tgttactggt cctgaacatc cagaattagg   9180
aagcttactg tggcctcaca ccacattaga atttgttctg gaaatcaagc tataggaaca   9240
agagatcgta agtcatatta tactatcaac ctaaattcca gtctgacaat tcctttgcaa   9300
aattgtgtaa aactccctta tattgctagt tgtaggaaaa acatagttat taaacctgat   9360
tcccaaacca taatctgtga aaattgtgga atgtttactt gcattgattt gacttttaat   9420
tggcagcacc gtattctact aggaagagca agagagggtg tgtggatcct tgtgtccatg   9480
gaccgaccat gggaggcttc gctatccatc catattttaa cggaagtatt aaaaggaatt   9540
ctaactagat ccaaaagatt catttttact ttgatggcag tgattatggg cctcattgca   9600
gtcacagcta ctgctgcggc tgctggaatt gctttacact cctctgttca aactgcagaa   9660
tacgtaaatg attggcaaaa gaattcctca aaattgtgga attctcagat ccaaatagat   9720
caaaaattgg caaaccaaat taatgatctt agacaaactg tcatttggat gggagaggct   9780
catgagcttg gaatatcttt ttcagttacg atgtgactgg aatacatcag atttttgtgt   9840
tacaccacaa gcctataatg agtctgagaca tcactgggac atggttagat gccatctgca   9900
aggaggagaa gataatctta ctttagacat ttcaaaatta aaagaatttt tttttctttt   9960
gagacagagt ctcgctctgt cgcccaggct ggagtgcagt ggcgtgatct cagctcactg   10020
caagttccgc ctcctgggtt tacaccattc tcctgcctca gcctcccaag tagttgggac   10080
tacaggagcc caccaccatg cctggctaat tttttttggg tttttaatag agatggagtt   10140
tcaccgtgtt agccaggatg gtctcgatct cctgaccttg tgatctgccc accttggcct   10200
cccaaagtgc tgggattaca gtcgtgagcc accgtgccca gccaagaaaa aatttttgag   10260
gcatcaaaag cccatttaaa tttggtgcca ggaacggaga caatcgtgaa agctgctgat   10320
agcctcacaa atcttaagcc agtcacttgg gttaaaagca tcagaagttt cactattgta   10380
aatttcatat taatccttgt atgcctgttc tgtctgttgt tagtctacag gtgtatccag   10440
cagctccaaa gagacagcaa ccagcaagaa tgggccatag tgacgatggt ggttttgtca   10500
aaaagaaaag gggggatat gtaaggaaaa gagagatcag actttcactg tgtctatgta   10560
gaaaggaag acataagaaa ctccattttg atctgtacta agaaaaattg ttttgccttg   10620
agatgctgtt aatctgtaac tttagcccca accctgtgct cacggaaaca tgtgctgtaa   10680
ggtttaaggg atctagggct gtgcaggatg taccttgtta acaatatgtt tgcaggcagt   10740
atgtttggta aaagtcatcg ccattctcca ttctcgatta accaggggct caatgcactg   10800
tggaaagcca caggaacctc tgcccaagaa agcctggctg ttgtgggaag tcagggaccc   10860
cgaatggagg gaccagctgg tgctgcatca ggaaacataa attgtgaaga tttcttggac   10920
atttatcagt ttccaaaatt aatactttta taatttctta cacctgtctt actttaatct   10980
cttaatcctg ttatctttgt aagctgagga tatacgtcac ctcaggacca ctattgtaca   11040
aattgattgt aaaacatgtt cacatgtgtt tgaacaatat gaaatcagtg caccttgaaa   11100
atgaacagaa taacagtgat tttagggaac aaaggaagac aaccataagg tctgactgcc   11160
```

```
tgaggggtcg ggcaaaaagc catatttttc ttcttgcaga gagcctataa atggacgtgc    11220
aagtaggaga gatattgcta aattctttc ctagcaagga atataatact aagaccctag    11280
ggaaagaatt gcattcctgg ggggaggtct ataaacggcc gctctgggag tgtctgtcct    11340
atgtggttga gataaggact gagatacgcc ctggtctcct gcagtaccct caggcttact    11400
aggattggga aaccccagtc ctggtaaatt tgaggtcagg ccggttcttt gctctgaacc    11460
ctgttttctg ttaagatgtt tatcaagaca atacatgcac cgctgaacat agacccttat    11520
caggagtttc tgattttgct ctggtcctgt ttcttcagaa gcatgtcatc tttgctctgc    11580
cttctgccct ttgaagcatg tgatctttgt gacctactcc ctgttcatac acccctcccc    11640
ttttaaaatc cctaataaaa acttgctggt tttgtggctc agggggggcat catggaccta    11700
ccaatacgtg atgtcacccc cggtggccca gctgtaaaat tcctttcttt atactcttat    11760
ttctcagacc agctgacact tagggaaaat agaaagaacc tatgttgaaa tattggaggc    11820
gggttccccc gatacctggg tattgtccaa ggtttccttt gctgaggagg attagtaaaa    11880
ggaatgcctc catctcctgc atgtccctgg gaacagaatg ttcccaccaa ccaccctgtg    11940
gctggaggcg ggatatgctg gcagcaatgc tgctctatta ctctttgcta cactgagatg    12000
tttgggtgga gagaagcata aatctggcct atgtgcacat ctgggcacag caccttcctt    12060
tgaacttatt tgtgacacag attcctttgc tcacgttttc ctgttgactt tctcaccact    12120
caccctattc tcctgtggca ttcgccttgc ggagatagtg aaaatagtaa taaatactga    12180
gggaactcag actgagggaa ctcagactgg gcagaccggg gccagtgtgg gtcctccata    12240
tgctgagcgc cggttccctg ggcccactgt tctttctcta tactttgtct ctgtgcctta    12300
ttttctcagt ctctcattcc acctgatgag aaatacccac aggtgtggag gggctggccc    12360
ccttca                                                                12366
```

<210> 76
<211> 2148
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 76
```
atggggcaaa ctgaaagtaa atatgcctct tatctcagct ttattaaaat tcttttaaga    60
agaggggggag ttagagcttc tacagaaaat ctaattacgc tatttcaaac aatagaacaa    120
ttctgcccat ggtttccaga acagggaact ttagatctaa aagattggga aaaaattggc    180
aaagaattaa aacaagcaaa tagggaaggt aaaatcatcc cacttacagt atggaatgat    240
tgggccatta ttaaagcaac tttagaacca tttcaaacag gagaagatat tgtttcagtt    300
tctgatgccc ctaaaagctg tgtaacagat tgtgaagaag aggcagggac agaatcccag    360
caaggaacgg aaagttcaca ttgtaaatat gtagcagagt ctgtaatggc tcagtcaacg    420
caaaatgttg actacagtca attacaggag ataatatacc ctgaatcatc aaaattgggg    480
gaaggaggtc cagaatcatt ggggccatca gagcctaaac cacgatcgcc atcaactcct    540
cctcccgtgg ttcagatgcc tgtaacatta caacctcaaa cgcaggttag acaagcacaa    600
accccaagag aaaatcaagt agaaagggac agagtctcta tcccggcaat gccaactcag    660
atacagtatc cacaatatca gccggtagaa aataagaccc aaccgctggt agtttatcaa    720
taccggctgc caaccgagct tcagtatcgg cctccttcag aggttcaata cagacctcaa    780
gcggtgtgtc ctgtgccaaa tagcacggca ccataccagc aacccacagc gatggcgtct    840
aattcaccag caacacagga cgcggcgctg tatcctcagc cgcccactgt gagacttaat    900
cctacagcat cacgtagtgg acagggtggt gcactgcatg cagtcattga tgaagccaga    960
aaacagggcg atcttgaggc atggcggttc ctggtaattt tacaactggt acaggccggg    1020
gaagagactc aagtaggagc gcctgcccga gctgagacta gatgtgaacc tttcaccatg    1080
aaaatgttaa aagatataaa ggaaggagtt aaacaatatg gatccaactc cccttatata    1140
agaacattat agattccat tgctcatgga aatagactta ctccttatga ctgggaaatt    1200
ttggccaaat cttcccttc atcctctcag tatctacagt ttaaaacctg gtggattgat    1260
ggagtacaag aacaggtacg aaaaaatcag gctactaagc ccactgttaa tatagacgca    1320
gaccaattgt taggaacagg tccaaattgg agcaccatta ccaacaatc agtgatgcag    1380
aatgaggcta ttgaacaagt aagggctatt tgcctcaggg cctggggaaa aattcaggac    1440
ccaggaacag cttttccctat taattcaatt agacaaggct ctaaagagcc atatcctgac    1500
tttgtggcaa gattacaaga tgctgctcaa aagtctatta cagatgacaa tgcccgaaaa    1560
gttattgtag aattaatggc ctatgaaaat gcaaatccag aatgtcagtc ggccataaag    1620
ccattaaaag gaaaagttcc agcaggagtt gatgtaatta cagaatatgt gaaggcttgt    1680
gatgggattg gaggagctat gcataaggca atgctaatgg ctcaagcaat gaggggggctc    1740
actctaggag acaagttag aacatttggg aaaaaatgtt ataattgtgg tcaaatcggt    1800
catctgaaaa ggagttgccc agtcttaaat aaacagaata taataaatca agctattaca    1860
gcaaaaaata aaaagccatc tggcctgtgt ccaaaatgtg aaaaggaaa cattgggcc    1920
aatcaatgtc attctaaatt tgataaggat gggcaaccat tgtcgggaaa caggaagagg    1980
ggccagcctc aggccccca acaaactggg gcattccag ttcaactgtt tgttcctcag    2040
ggttttcaag acaacaacc cctacagaaa ataccaccac ttcagggagt cagccaatta    2100
```

caacaatcca acagctgtcc cgcgccacag caggcagcac cgcagtaa                  2148

<210>    77
<211>    2151
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Manipulated Gag

<400>    77
atgggccaga ccgagagcaa gtacgccagc tacctgagct tcatcaagat cctgctgcgc    60
cgcggcggcg tgcgcgccag caccgagaac ctgatcaccc tgttccagac catcgagcag    120
ttctgccccct ggttccccga gcagggcacc ctggacctga aggactggga gaagatcggc    180
aaggagctga agcaggccaa ccgcgagggc aagatcatcc ccctgaccgt gtggaacgac    240
tgggccatca tcaaggccac cctggagccc ttccagaccg gcgaggacat cgtgagcgtg    300
agcgacgccc ccaagagctg cgtgaccgac tgcgaggagg aggccggcac cgagagccag    360
cagggcaccg agagcagcca ctgcaagtac gtggccgaga gcgtgatggc ccagagcacc    420
cagaacgtgg actacagcca gctgcaggag atcatctacc ccgagagcag caagctgggc    480
gagggcggcg ccgagagcct gggccccagc gagcccaagc cccgcagccc cagcaccccc    540
cccccgtgg tgcagatgcc cgtgaccctg cagccccaga cccaggtgcg ccaggcccag    600
accccccgcg agaaccaggt ggagcgcgac cgcgtgagca tccccgccat gcccacccag    660
atccagtacc cccagtacca gcccgtggag aacaagaccc agcccctggt ggtgtaccag    720
taccgcctgc ccaccgagct gcagtaccgc ccccccagcg aggtgcagta ccgcccccag    780
gccgtgtgcc ccgtgcccaa cagcaccgcc ccctaccagc agcccaccgc catggccagc    840
aacagccccg ccacccagga cgccgccctg taccccccagc cccccaccgt gcgcctgaac    900
cccaccgcca gccgcagcgg ccagggcggc gccctgcacg ccgtgatcga cgaggcccgc    960
aagcagggcg acctggaggc ctggcgcttc ctggtgatcc tgcagctggt gcaggccggc    1020
gaggagaccc aggtgggcgc ccccgcccgc gccgagaccc gctgcgagcc cttcaccatg    1080
aagatgctga aggacatcaa ggagggcgtg aagcagtacg cagcaacag ccccctacatc    1140
cgcaccctgc tggacagcat cgcccacggc aaccgcctga ccccctacga ctgggagatc    1200
ctggccaaga gcagcctgag cagcagccag tacctgcagt tcaagacctg gtggatcgac    1260
ggcgtgcagg agcaggtgcg caagaaccag gccaccaagc ccaccgtgaa catcgacgcc    1320
gaccagctgc tgggcaccgg ccccaactgg agcaccatca accagcagag cgtgatgcag    1380
aacgaggcca tcgagcaggt gcgcgccatc tgcctgcgcg cctggggcaa gatccaggac    1440
cccggcaccg ccttccccat caacagcatc gccagggca gcaaggagcc ctaccccgac    1500
ttcgtggccc gcctgcagga cgccgcccag aagagcatca ccgacgacaa cgcccgcaag    1560
gtgatcgtgg agctgatggc ctacgagaac gccaacccccg agtgccagag cgccatcaag    1620
cccctgaagg gcaaggtgcc cgccggcgtg acgtgatca ccgagtacgt gaaggcctgc    1680
gacggcatcg gcggcgccat gcacaaggcc atgctgatgg cccaggccat gcgcggcctg    1740
accctgggcg gccaggtgcg caccttcggc aagaagtgct acaactcgg ccagatcggc    1800
cacctgaagc gcagcgtgcc cgtgctgaac aagcagaaca tcatcaacca ggccatcacc    1860
gccaagaaca agaagcccag cggcctgtgc cccaagtgcg caagggcaa cgactgggcc    1920
aaccagtgcc acagcaagtt cgacaaggac ggccagcccc tgagcggcaa ccgcaagcgc    1980
ggccagcccc aggccccca gcagaccggc gccttccccg tgcagctgtt cgtgccccag    2040
ggcttccagg gccagcagcc cctgcagaag atcccccccc tgcagggcgt gagccagctg    2100
cagcagagca cagctgccc cgcccccag caggccgccc cccaggctta a    2151

<210>    78
<211>    715
<212>    PRT
<213>    Human endogenous retrovirus, K family (HERV-K)

<400>    78
Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1               5                   10                  15

Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile
            20                  25                  30

Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
            35                  40                  45

Page 82

Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
    50              55                  60

Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65              70              75                  80

Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
                85              90              95

Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
            100             105             110

Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
        115             120             125

Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
    130             135             140

Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145             150             155             160

Glu Gly Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
            165             170             175

Pro Ser Thr Pro Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
            180             185             190

Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
        195             200             205

Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
    210             215             220

Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225             230             235             240

Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
            245             250             255

Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
            260             265             270

Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
        275             280             285

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
    290             295             300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305             310             315             320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
            325             330             335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
        340             345             350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
    355             360             365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
    370             375             380

Page 83

```
Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
385             390             395             400

Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
            405             410             415

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
            420             425             430

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
        435             440             445

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
    450             455             460

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
465             470             475             480

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
            485             490             495

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
            500             505             510

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
    515             520             525

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
    530             535             540

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
545             550             555             560

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
            565             570             575

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
            580             585             590

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
            595             600             605

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
    610             615             620

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
625             630             635             640

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
            645             650             655

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
            660             665             670

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
            675             680             685

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
    690             695             700

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
705             710             715
```

```
<210>    79
<211>    716
<212>    PRT
<213>    Artificial Sequence

<220>
<223>    Manipulated Gag

<400>    79
Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1               5                   10                  15

Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile
            20                  25                  30

Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
        35                  40                  45

Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
    50                  55                  60

Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65                  70                  75                  80

Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
                85                  90                  95

Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
            100                 105                 110

Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
        115                 120                 125

Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
    130                 135                 140

Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145                 150                 155                 160

Glu Gly Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
                165                 170                 175

Pro Ser Thr Pro Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
            180                 185                 190

Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
        195                 200                 205

Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
    210                 215                 220

Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225                 230                 235                 240

Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
                245                 250                 255

Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
            260                 265                 270

Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
        275                 280                 285
```

Page 85

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
    290             295             300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305             310             315             320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
            325             330             335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
            340             345             350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
            355             360             365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
    370             375             380

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
385             390             395             400

Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
            405             410             415

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
            420             425             430

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
            435             440             445

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
    450             455             460

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
465             470             475             480

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
            485             490             495

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
            500             505             510

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
            515             520             525

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
    530             535             540

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
545             550             555             560

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
            565             570             575

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
            580             585             590

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
            595             600             605

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
    610             615             620

Page 86

```
Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
625             630             635             640

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
                645             650             655

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
            660             665             670

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
        675             680             685

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
    690             695             700

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln Ala
705             710             715
```

<210> 80
<211> 6486
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.gagopt PCAV vector

<400> 80
```
gccgcggaat tcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat   60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta   120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata   180
acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat   240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga   300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc   360
ccctattgac gtcaatgacg taaatggcc cgcctggcat tatgcccagt acatgacctt   420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat   480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag   540
tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc   600
aaaatgtcgt aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga   660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg   720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg   780
cattggaacg cggattcccc gtgccaagat gacgtaagt accgcctata gactctatag   840
gcacacccct ttggctctta tgcatgctat actgttttg gcttggggc tatacacccc   900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat   960
tattgaccac tcccctattg gtgacgatac tttccattac taatccataa catggctctt   1020
tgccacaact atctctattg gctatatgcc aatactctgt ccttcagaga ctgacacgga   1080
ctctgtattt ttacaggatg gggtcccatt tattatttac aaattcacat atacaacaac   1140
gccgtccccc gtgcccgcag ttttattaa acatagcgtg ggatctccac gcgaatctcg   1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc   1260
tggtcccatg cctccagcgg ctcatggtcg ctcggcagct ccttgctcct aacagtggag   1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg   1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga   1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag   1500
tcagaggtaa ctcccgttgc ggtgctgtta acggtggagg gcagtgtagt ctgagcagta   1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt   1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgggccaga ccgagagcaa   1680
gtacgccagc tacctgagct tcatcaagat cctgctgcgc cgcggcggcg tgcgcgccag   1740
caccgagaac ctgatcaccc tgttccagac catcgagcag ttctgcccct ggttccccga   1800
gcagggcacc ctggacctga aggactggga agatcggc aaggagctga agcaggccaa   1860
ccgcgagggc aagatcatcc ccctgaccgt gtggaacgac tgggccatca tcaaggccac   1920
cctggagccc ttccagaccg gcgaggacat cgtgagcgtg agcgacgccc ccaagagctg   1980
cgtgaccgac tgcgaggagg aggccggcac cgagagccag cagggcaccg agagcagcca   2040
ctgcaagtac gtggccgaga gcgtgatggc ccagagcacc cagaacgtgg actacagcca   2100
gctgcaggag atcatctacc ccgagagcag caagctgggc gagggcggcc ccgagagcct   2160
```
Page 87

```
gggccccagc gagcccaagc cccgcagccc cagcacccc ccccccgtgg tgcagatgcc    2220
cgtgaccctg cagccccaga cccaggtgcg ccaggcccag accccccgcg agaaccaggt    2280
ggagcgcgac cgcgtgagca tccccgccat gcccacccag atccagtacc cccagtacca    2340
gcccgtggag aacaagaccc agcccctggt ggtgtaccag taccgcctgc ccaccgagct    2400
gcagtaccgc ccccccagcg aggtgcagta ccgcccccag gccgtgtgcc ccgtgcccaa    2460
cagcaccgcc ccctaccagc agcccaccgc catggccagc aacagccccg ccacccagga    2520
cgccgccctg tacccccagc ccccaccgt gcgcctgaac cccaccgcca gccgcagcgg    2580
ccagggcggc gccctgcacg ccgtgatcga cgaggccgc aagcagggcg acctggaggc    2640
ctggcgcttc ctggtgatcc tgcagctggt gcaggccggc gaggagaccc aggtgggcgc    2700
ccccgcccgc gccgagaccc gctgcgagcc cttcaccatg aagatgctga aggacatcaa    2760
ggagggcgtg aagcagtacg gcagcaacag cccctacatc cgcaccctgc tggacagcat    2820
cgcccacggc aaccgcctga cccctacga ctgggagatc ctggccaaga gcagcctgag    2880
cagcagccag tacctgcagt tcaagacctg gtggatcgac ggcgtgcagg agcaggtgcg    2940
caagaaccag gccaccaagc ccaccgtgaa catcgacgcc gaccagctgc tgggcaccgg    3000
ccccaactgg agcaccatca accagcagag cgtgatgcag aacgaggcca tcgagcaggt    3060
gcgcgccatc tgcctgcgcg cctggggcaa gatccaggac cccggcaccg ccttccccat    3120
caacagcatc cgccagggca gcaaggagcc ctaccccgac ttcgtggccc gcctgcagga    3180
cgccgcccag aagagcatca ccgaccgacaa cgcccgcaag gtgatcgtgg agctgatggc    3240
ctacgagaac gccaaccccg agtgccagag cgccatcaag cccctgaagg gcaaggtgcc    3300
cgccggcgtg gacgtgatca ccgagtacgt gaaggcctgc gacggcatcg gcggcgccat    3360
gcacaaggcc atgctgatgg cccaggccat gcgcggcctg accctgggcg ccaggtgcg    3420
caccttcggc aagaagtgct acaactgcgg ccagatcggc cacctgaagc gcagctgccc    3480
cgtgctgaac aagcagaaca tcatcaacca ggccatcacc gccaagaaca gaagcccag    3540
cggcctgtgc cccaagtgcg gcaagggcaa gcactgggcc aaccagtgcc acagcaagtt    3600
cgacaaggac ggccagcccc tgagcggcaa ccgcaagcgc ggccagcccc aggccccca    3660
gcagaccggc gccttcccg tgcagctgtt cgtgccccag ggcttccagg ccagcagcc    3720
cctgcagaag atccccccc tgcagggcgt gagccagctg cagcagagca cagctgccc    3780
cgcccccag caggccgccc cccaggctta agaattcaga ctcgagcaag tctagaaagc    3840
catggatatc ggatccacta cgcgttagag ctcgctgatc agcctcgact gtgccttcta    3900
gttgccagcc atctgttgtt tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca    3960
ctcccactgt cctttcctaa taaaatgagg aaattgcatc gcattgtctg agtaggtgtc    4020
attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata    4080
gcagggggt gggcgaagaa ctccagcatg agatccccgc gctggaggat catccagccg    4140
gcgtcccgga aaacgattcc gaagcccaac ctttcataga aggcggcggt ggaatcgaaa    4200
tctcgtgatg gcaggttggg cgtcgcttgg tcggtcattt cgaaccccag agtcccgctc    4260
agaagaactc gtcaagaagg cgatagaagg cgatgcgctg cgaatcggga gcggcgatac    4320
cgtaaagcac gaggaagcgg tcagcccatt cgccgccaag ctcttcagca atatcacggg    4380
tagccaacgc tatgtcctga tagcggtccg ccacacccag ccggccacag tcgatgaatc    4440
cagaaaagcg gccattttcc accatgatat tcggcaagca ggcatcgcca tgggtcacga    4500
cgagatcctc gccgtcgggc atgcgcgcct tgagcctggc gaacagttcg gctggcgcga    4560
gcccctgatg ctcttcgtcc agatcatcct gatcgacaag accggcttcc atccgagtac    4620
gtgctcgctc gatgcgatgt ttcgcttggt ggtcgaatgg gcaggtagcc ggatcaagcg    4680
tatgcagccg ccgcattgca tcagccatga tggatacttt ctcggcagga gcaaggtgag    4740
atgacaggag atcctgcccc ggcacttcgc ccaatagcag ccagtccctt cccgcttcag    4800
tgacaacgtc gagcacagct gcgcaaggaa cgcccgtcgt ggccagccac gatagccgcg    4860
ctgcctcgtc ctgcagttca ttcagggcac cggacaggtc ggtcttgaca aaaagaaccg    4920
ggcgcccctg cgctgacagc cggaacacgg cggcatcaga gcagccgatt gtctgttgtg    4980
cccagtcata gccgaatagc ctctccaccc aagcggccgg agaacctgcg tgcaatccat    5040
cttgttcaat catgcgaaac gatcctcatc ctgtctcttg atcagatctt gatcccctgc    5100
gccatcagat ccttggcggc aagaaagcca tccagtttac tttgcagggc ttcccaacct    5160
taccagaggg cgccccagct ggcaattccg gttcgcttgc tgtccataaa accgcccagt    5220
ctagctatcg ccatgtaagc ccactgcaag ctacctgctt tctctttgcg cttgcgtttt    5280
cccttgtcca gatagcccag tagctgacat tcatccgggg tcagcaccgt ttctgcggac    5340
tggctttcta cgtgttccgc ttcctttagc agcccttgcg ccctgagtgc ttgcggcagc    5400
gtgaagctaa ttcatggtta aatttttgtt aaatcagctc atttttttaac caataggccg    5460
aaatcggcaa aatcccttat aaatcaaaag aatagcccga gatagggttg agtgttgttc    5520
cagtttggaa caagagtcca ctattaaaga cgtggactcc aacgtcaaa gggcgaaaaa    5580
ccgtctatca gggcgatggc cggatcagct tatgcggtgt gaaataccgc acagatgcgt    5640
aaggagaaaa taccgcatca ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc    5700
ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag cggtaatac ggttatccac    5760
agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa    5820
ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca    5880
caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc    5940
```

```
gtttcccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata    6000
cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta    6060
tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca    6120
gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga    6180
cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg    6240
tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg    6300
tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg    6360
caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag    6420
aaaaaaagga tctcaagaag atcctttgat cttttctact gaacggtgat ccccaccgga    6480
attgcg                                                                6486
```

<210> 81
<211> 2103
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 81

```
atgaacccaa gcgagatgca aagaaaagca cctccgcgga gacggagaca tcgcaatcga    60
gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag aagaacagat gaagttgcca    120
tccaccaaga aggcagagcc gccaacttgg gcacaactaa agaagctgac gcagttagct    180
acaaaatatc tagagaacac aaaggtgaca caaaccccag agagtatgct gcttgcagcc    240
ttgatgattg tatcaatggt ggtaagtctc cctatgcctg caggagcagc tgcagctaac    300
tatacctact gggcctatgt gcctttcccg cccttaattc gggcagtcac atggatggat    360
aatcctacag aagtatatgt taatgatagt gtatgggtac ctggccccat agatgatcgc    420
tgccctgcca aacctgagga agaagggatg atgataaata tttccattgg gtatcattat    480
cctcctattt gcctagggag agcaccagga tgtttaatgc ctgcagtcca aaattggttg    540
gtagaagtac ctactgtcag tcccatctgt agattcactt atcacatggt aagcgggatg    600
tcactcaggc cacgggtaaa ttatttacaa gacttttctt atcaaagatc attaaaattt    660
agacctaaag ggaaaccttg ccccaaggaa attcccaaag aatcaaaaaa tacagaagtt    720
ttagtttggg aagaatgtgt ggccaatagt gcggtgatat tacaaaacaa tgaattcgga    780
actattatag attgggcacc tcgaggtcaa ttctaccaca attgctcagg acaaactcag    840
tcgtgtccaa gtgcacaagt gagtccagct gttgatagcg acttaacaga aagtttagac    900
aaacataagc ataaaaaatt gcagtctttc tacccttggg aatggggaga aaaaggaatc    960
tctaccccaa gaccaaaaat agtaagtcct gtttctggtc ctgaacatcc agaattatgg    1020
aggcttactg tggcttcaca ccacattaga atttggtctg gaaatcaaac tttagaaaca    1080
agagatcgta agccatttta tactattgac ctgaattcca gtctaacagt tcctttacaa    1140
agttgcgtaa agccccctta tatgctagtt gtaggaaata tagttattaa accagactcc    1200
cagactataa cctgtgaaaa ttgtagattg cttacttgca ttgattcaac ttttaattgg    1260
caacaccgta ttctgctggt gagagcaaga gagggcgtgt ggatccctgt gtccatggac    1320
cgaccgtggg aggcctcgcc atccgtccat attttgactg aagtattaaa aggtgtttta    1380
aatagatcca aaagattcat ttttacttta attgcagtga ttatgggatt aattgcagtc    1440
acagctacgg ctgctgtagc aggagttgca ttgcactctt ctgttcagtc agtaaacttt    1500
gttaatgatt ggcaaaaaaa ttctacagga ttgtggaatt cacaatctag tattgatcaa    1560
aaattggcaa atcaaattaa tgatcttaga caaactgtca tttggatggg agacagactc    1620
atgagcttag aacatcgttt ccagttacaa tgtgactgga atacgtcaga tttttgtatt    1680
acaccccaaa tttataatga gtctgagcat cactgggaca tggttagacg ccatctacag    1740
ggaagagaag ataatctcac tttagacatt tccaaattaa aagaacaaat tttcgaagca    1800
tcaaaagccc atttaaattt ggtgccagga actgaggcaa ttgcaggagt tgctgatggc    1860
ctcgcaaatc ttaaccctgt cacttgggtt aagaccattg gaagtactac gattataaat    1920
ctcatattaa tccttgtgtg cctgtttttgt ctgttgttag tctgcaggtg tacccaacag    1980
ctccgaagag acagcgacca tcgagaacgg gccatgatga cgatggcggt tttgtcgaaa    2040
agaaaagggg gaaatgtggg gaaaagcaag agagatcaga ttgttactgt gtctgtggcc    2100
taa                                                                   2103
```

<210> 82
<211> 2103
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified env sequence

<400> 82

```
atgaacccca gcgagatgca gcgcaaggcc ccccccgcc gccgccgcca ccgcaaccgc    60
gcccccctga cccacaagat gaacaagatg gtgaccagcg aggagcagat gaagctgccc   120
agcaccaaga aggccgagcc ccccacctgg gcccagctga agaagctgac ccagctggcc   180
accaagtacc tggagaacac caaggtgacc cagacccccg agagcatgct gctggccgcc   240
ctgatgatcg tgagcatggt ggtgagcctg cccatgcccg ccggcgccgc cgccgccaac   300
tacacctact gggcctacgt gcccttcccc cccctgatcc gcgccgtgac ctggatggac   360
aaccccaccg aggtgtacgt gaacgacagc gtgtgggtgc ccggccccat cgacgaccgc   420
tgccccgcca gcccgagga ggagggcatg atgatcaaca tcagcatcgg ctaccactac   480
ccccccatct gcctgggccg cgccccggc tgcctgatgc cgccgtgca gaactggctg   540
gtggaggtgc ccaccgtgag ccccatctgc cgcttcacct accacatggt gagcggcatg   600
agcctgcgcc ccgcgtgaa ctacctgcag gacttcagct accagcgcag cctgaagttc   660
cgccccaagg gcaagccctg ccccaaggag atccccaagg agagcaagaa caccgaggtg   720
ctggtgtggg aggagtgcgt ggccaacagc gccgtgatcc tgcagaacaa cgagttcggc   780
accatcatcg actgggcccc ccgcggccag ttctaccaca actgcagcgg ccagacccag   840
agctgcccca gcgcccaggt gagccccgcc gtggacagcg acctgaccga gagcctggac   900
aagcacaagc acaagaagct gcagagcttc taccctggg agtggggcga aagggcatc    960
agcacccccc gccccaagat cgtgagcccc gtgagcggcc ccgagcaccc cgagctgtgg   1020
cgcctgaccg tggccagcca ccacatccgc atctggagcg gcaaccagac cctggagacc   1080
cgcgaccgac agcccttcta caccatcgac ctgaacagca gcctgaccgt gcccctgcag   1140
agctgcgtga gcccccccta catgctggtg gtgggcaaca tcgtgaacaa gcccgacagc   1200
cagaccatca cctgcgagaa ctgcccgcca ctgacctgca tcgacagcac cttcaactgg   1260
cagcaccgca tcctgctggt gcgcgcccgc gagggcgtgt ggatccccgt gagcatggac   1320
cgcccctggg aggccagccc cagcgtgcac atcctgaccg aggtgctgaa gggcgtgctg   1380
aaccgcagca agcgcttcat cttcacccctg atcgccgtga tcatgggcct gatcgccgtg   1440
accgccaccg ccgccgtggc cggcgtggcc ctgcacagca gcgtgcagag cgtgaacttc   1500
gtgaacgact ggcagaagaa cagcacccgc ctgtggaaca gccagagcag catcgaccag   1560
aagctggcca accagatcaa cgacctgcgc cagaccgtga tctggatggg cgaccgcctg   1620
atgagcctgg agcaccgctt ccagctgcag tgcgactgga caccagcga cttctgcatc   1680
acccccaga tctacaacga gagcgagcac cactgggaca tggtgcgccg ccacctgcag   1740
ggccgcgagg acaacctgac cctggacatc agcaagctga aggagcagat cttcgaggcc   1800
agcaaggccc acctgaacct ggtgcccggc accgaggcca tcgccggcgt ggccgacggc   1860
ctggccaacc tgaacccccgt gacctgggtg aagaccatcg gcagcaccac catcatcaac   1920
ctgatcctga tcctggtgtg cctgttctgc ctgctgctgg tgtgccgctg cacccagcag   1980
ctgcgccgcg acagcgacca ccgcgagcgc gccatgatga ccatggccgt gctgagcaag   2040
cgcaagggcg gcaacgtggg caagagcaag cgcgaccaga tcgtgaccgt gagcgtggcc   2100
taa                                                                 2103
```

<210>  83
<211>  700
<212>  PRT
<213>  Human endogenous retrovirus, K family (HERV-K)

<400>  83

```
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1               5                   10                  15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
            20                  25                  30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
        35                  40                  45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
    50                  55                  60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65                  70                  75                  80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
                85                  90                  95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
            100                 105                 110
```

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
        115                 120                 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
        130             135                 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145                 150                 155                 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
                165             170                 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
            180                 185                 190

Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
        195                 200                 205

Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
    210                 215                 220

Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
225                 230                 235                 240

Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
                245                 250                 255

Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
            260                 265                 270

His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
        275                 280                 285

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
    290                 295                 300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305                 310                 315                 320

Ser Thr Pro Arg Pro Lys Ile Val Ser Pro Val Ser Gly Pro Glu His
                325                 330                 335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
            340                 345                 350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
        355                 360                 365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
    370                 375                 380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385                 390                 395                 400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
                405                 410                 415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
            420                 425                 430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
        435                 440                 445

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
    450             455             460

Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
465             470             475             480

Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
            485             490             495

Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
        500             505             510

Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
        515             520             525

Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
    530             535             540

His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
545             550             555             560

Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
            565             570             575

Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
        580             585             590

Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
        595             600             605

Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
    610             615             620

Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
625             630             635             640

Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
            645             650             655

Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
            660             665             670

Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
    675             680             685

Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val Ala
    690             695             700